Multi-State Alternate Assessment

2023 Science Technical Report

3/4/24

Prepared by Cognia for MSAA

TABLE OF CONTENTS

Appendix References by Chapter

Chapter	1	2	3	4	5	6	7	8	9	10	11
Appendix		A, B, C	D, E, F	B		G	H	I	J, K, L, M	N, O, P	I, Q

Chapter 1. Overview of MSAA and 2023 Updates

1.1 Purposes and Uses of the MSAA and MSAA Science

The Multi-State Alternate Assessment (the MSAA) is a multidimensional, summative assessment system designed to promote increasingly higher academic outcomes for students with the most significant cognitive disabilities to prepare them for a broader array of post-secondary outcomes. The MSAA Science is designed to measure grade-level academic science content that is aligned with, and derived from, MSAA Partners' science content standards. This test contains built-in supports that do not change the construct being measured, to allow students to use materials they are most familiar with and communicate what they know and what they can do as independently as they are able. The MSAA Science, for some of the MSAA Partners, is part of the comprehensive assessment program, which includes the MSAA ELA and Mathematics. Many aspects of the assessment program delivery for the MSAA Science are consistent with what is provided for MSAA ELA and Mathematics, including the online assessment platform, training requirements, and participation guidelines. One difference in science is that all three grades are linear or fixed form tests instead of stage-adaptive assessments.

The MSAA Science is an alternate assessment aligned with alternate academic achievement standards (AA-AAAS) as described in the Elementary and Secondary Education Act (ESEA). This law mandates that all students participate in assessments that measure student achievement of grade-level content standards. The MSAA Science was developed to ensure that all students with the most significant cognitive disabilities can participate in a summative assessment that provides a measure of what they know and can do in relation to grade-level science content standards. To ensure that MSAA Science measures student achievement of alternate academic achievement standards aligned to grade-level science content standards, this technical report provides the standard psychometric analyses and descriptions of technical procedures commonly found in state assessment technical reports.

1.2 Initial Administration of the MSAA Science

The 2022 administration of the MSAA Science was the first operational administration. The MSAA Science Partners for 2022 were Arizona, Bureau of Indian Education (BIE), Maine, and the U.S. Virgin Islands. The impact of COVID-19 worldwide resulted in cancellation of the 2020 field test administration (which was scheduled to occur in Maine only) and continued to influence many MSAA Partners' participation levels in 2021 and 2022. The impact for the MSAA Science was that the operational implementation was delayed from 2021 to 2022. The MSAA Science was administered to Arizona and Maine only as a census field test in 2021. A standard setting was conducted in July 2022 for the MSAA Science, resulting in provisional performance levels. The performance levels were reviewed and finalized via another standard setting convening in summer 2023. The standard-setting technical report is available at www.msaastates.com. Additional detailed information about work leading up to the 2022 administration is available in Chapter 2.

In fall 2021, in preparation for the census field-test administration in 2021, MSAA Science sample items were available through the online assessment platform for grades 5 and High School (HS), which both included one standalone set (three items) per grade. In fall 2022, the number of MSAA Science sample items was expanded to include one standalone set and one cluster for each grade.

1.3 Current Year Updates

The MSAA Science Partners for 2023 were American Samoa, Arizona, Commonwealth of the Northern Mariana Islands (CNMI), Guam, Bureau of Indian Education (BIE), Maine, the U.S. Virgin Islands, and Vermont.

In fall 2023, additional sample items, as well as the previously available sample items, were available to teachers through the online assessment platform. Grades 5, 8, and HS now have sample tests inclusive of two standalone sets and two clusters, totaling 18 sample items per grade. The sample items include the grade-specific Directions for Test Administration (DTA), which correspond with the items in the online system to emulate and standardize the student testing experience. These sample items are located at https://www.msaaassessment.org/sample-items and can be accessed year-round. Additional detailed information about test design and items is available in Chapter 3.

Additionally, for the 2023 administration, test documentation was updated to reflect changes in the Test Administration Manual (TAM), MSAA Online Assessment System User Guide for Test Administrators, MSAA Online Assessment System User Guide for Test Coordinators, Directions for Test Administration (DTA), and the *MSAA 2023 Guide for Score Report Interpretation Guide*. The TAM, MSAA Online Assessment System User Guide for Test Administrators, MSAA Online Assessment System User Guide for Test Coordinators, DTA, and online training modules were revised from the previous version to streamline information and provide more clarity to Test Administrators (TAs) and Test Coordinators (TCs). Additional detailed information about this process is available in Chapter 5.

1.4 Intended MSAA Science Score Interpretations and Uses

The MSAA Science is designed, developed, and implemented to support four primary intended score interpretations and uses, described in the following sections.

Primary Intended MSAA Science Score Interpretation

The MSAA Science scores provide reliable and valid information about important knowledge and skills in elementary, middle, and high school multidimensional science concepts that students with the most significant cognitive disabilities are attaining.

Primary Intended MSAA Science Score Uses

- Schools and districts use the MSAA Science and its results to monitor trends in student performance and design professional development for teachers on how to monitor trends.
- The MSAA Science and its results are used to help teachers integrate MSAA scores and other information into their instructional planning.
- Parents understand and interpret MSAA Science scores and other information correctly to understand what their child knows and can do.

The intended score interpretation and uses stated here align with the claim developed for the science assessment. The claim states:

- Students can use the majority[1] of the disciplinary core ideas, practices, and crosscutting concepts as stated in the grade-level Extended Performance Expectations (EPEs) to address moderately complex science phenomena and problems, some concrete and some abstract.

[1] *Majority is intended as at least half of the science content.*

The assumptions that underlie the intended interpretations and uses of MSAA Science scores, and a summary of the evidence that supports these assumptions, are presented in Chapter 11.

1.5 Validity Arguments for the MSAA

The 2023 technical report describes several procedural and psychometric processes of the MSAA program. These processes contribute to the accumulation of validity evidence to support MSAA score interpretations and uses. This report presents documentation to substantiate the intended interpretations and uses of MSAA test scores (AERA et al., 2014). Each section in this report contributes important information about the MSAA tests: test design and development, test alignment, test administration, scoring, reliability, performance levels, and reporting. The evidence available to support validity arguments for intended MSAA test score interpretations and uses is summarized in Chapter 11. The high-level validity arguments are listed below.

The phrase "intended score interpretations for uses" appears several times in the *Standards for Educational and Psychological Testing* (AERA et al., 2014) and is the core of the field's views on validity and validation. For the MSAA and other assessment programs, the phrase refers broadly to **test scores** (e.g., total test scale scores, aggregations of test scores, the percentage of students at or above Level 3), and other test performance information elements (e.g., the definition of Level 3 in the performance level descriptors). The *Standards for Educational and Psychological Testing* provides 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 overall validity of score interpretations and uses. Moreover, these sources represent only a partial list of sources of evidence from the MSAA design, development, test administration, analysis, and reporting processes that are relevant to the overall validity arguments for intended interpretations and uses of MSAA scores and other information. Descriptions of the test development and review process and results from operational psychometric analyses (e.g., test forms equating) are other examples.

Primary Intended Score Interpretation: The MSAA Science scores provide reliable and valid information about important knowledge and skills in elementary, middle, and high school multidimensional science concepts that students with the most significant cognitive disabilities are attaining.
1.1 The content of the test represents the content of the standards (i.e., the Extended Performance Expectations).
1.2 MSAA Science test items are construct relevant. The elements related to this assumption involve the skills and cognitive processes needed to respond to a specific item, and their alignment with those in the PLDs.
1.3 Test administrations in MSAA states followed prescribed, standardized procedural requirements.
1.4 Test scores on the MSAA Science provide reliable information about student performance and accurate classifications into performance levels.
1.5 Item and test scoring were implemented accurately.
1.6 MSAA Science scores correlate with external indicators of student proficiency (i.e., concurrent, and predictive evidence).

Primary Intended Score Use 1: Schools and districts use the MSAA Science and its results to (a) monitor trends in school performance, and (b) design professional development for teachers on how to monitor trends.
2.1 Schools and districts use the MSAA Science and its results to monitor trends in school performance.

2.2 MSAA Science results are used to design professional development for teachers.

***Primary Intended Score Use 2*:** The MSAA Science and its results are used to help teachers integrate MSAA Science scores and other information with their instructional planning.
3.1 Teachers use the MSAA Science and its results to better integrate assessment with their instructional planning.
3.2 Teachers use MSAA Science scores and other information for instructional planning.

***Primary Intended Score Use 3*:** Parents understand and interpret MSAA Science scores and other information correctly to understand what their child knows and can do.
4.1 Parents find MSAA Science scores and other information useful for understanding what their child knows and can do.
4.2 Parents find MSAA Science scores and other information useful for understanding their child’s progress over time.

Chapter 2. History of the MSAA Science

The MSAA Science assesses grades 5, 8, and HS and is aligned with the state/entity content standards and the content from *A Framework for K-12 Science Education* (National Research Council, 2012). The MSAA Science is a computer-based, on-demand assessment, consisting solely of selected-response items. The items are written at distinct levels of complexity, representing different levels of skill and knowledge acquisition by students. The MSAA Science consists of 39 operational selected-response items and 9 field-test selected-response items per grade. The assessment is configured in two test sessions.

Students with the most significant cognitive disabilities often need materials and instructional strategies that are substantially adapted and scaffolded, providing built-in supports to meet their individual needs. When students begin to learn a new skill or acquire new knowledge, they need more support and scaffolding. Generally, as ongoing instruction is provided and students learn and develop mastery of that skill or knowledge, they should need less support.

The MSAA Science levels of complexity are designed to follow instructional practices as noted above where appropriate. The test items are developed with many scaffolds and supports embedded within the items. Students are provided additional support based on their individual requirements. The scaffolds and supports embedded with the items are not intended to change the construct of the science concepts being assessed.

The MSAA Science is designed to be administered one-on-one, delivered in an online format or via a paper-pencil/hybrid format as an accommodation if appropriate. The needs of the student are also addressed through other supports, such as assessment features built into the platform and accommodations such as using assistive technology, a scribe, and/or sign language. The allowable accommodations and their corresponding protocols and guidelines are explained in detail in the Test Administration Manual (TAM) and must be adhered to as stated. Accommodations used during testing must also be included in the student's IEP prior to testing. Appendix A contains the 2023 summary of accommodation usage frequencies for the MSAA. Test Administrators (TAs) have substantial leeway in developing a testing schedule, including the ability to start and stop a test depending on the engagement of the student.

2.1 Core Beliefs

The MSAA Science Partners believe that accessibility is central to the validity argument of the assessment, and that access to science content based on rigorous college- and career-ready academic standards is essential for a student to demonstrate what they know and can do, which leads to greater post-secondary outcomes. The original design claim highlights the high expectations that are a part of the MSAA Science. The MSAA Science design was informed by multiple stakeholder reviews to ensure inclusive accessibility and appropriately high expectations for learning. The comprehensive MSAA program was based on the same model of learning as was reflected in classroom resources and contains built-in supports and features that are appropriate to the student population that takes the assessments. In addition, MSAA Science Partners provide resources for intervention in communicative competence to ensure that all students have a way first to learn the concepts and then to show what they know and can do on the assessment.

2.2 Pre-administration Development

Work leading up to the 2023 MSAA Science operational administration began in late 2015. Cognia content and accessibility specialists (referred to as the Cognia development team) drove the work. Throughout the assessment design and development process, additional Cognia experts in content and accessibility (external to the Cognia development team on MSAA Science) as well as MSAA Science Partners and their state/entity educators provided various stakeholder reviews. The following Table 2-1 provides a summary of the work on the MSAA Science from the beginning through the first operational administration in 2022.

Table 2-1. Timeline for the MSAA Science Design through Administration

Activity	Purpose & Stakeholders Involved	Timeframe*
Claim, test design, and PLD development	Develop overarching claim that defines the outcome we expect from student performance and draft item and test design structure to elicit the desired claim. Develop policy level performance level descriptors (PLD). Completed by Cognia development team	Winter 2015
Priority general education PE selection and EPE development	Select prioritized general education Performance Expectations (PEs) for coverage of science concepts across elementary, middle, and high school levels. Develop test blueprints for each grade level. Develop EPEs at various levels of complexity for the prioritized PEs. Completed by Cognia development team	Spring–Winter 2016
Stakeholder review: test blueprints, design, and draft EPEs	Convene stakeholder committee to review the test blueprints, the test design, and draft version of the EPEs. Gather feedback to refine the assessment rationale and update the EPEs. Review completed by Cognia content and accessibility specialists external to MSAA Science team	December 2016
Refine and finalize EPEs and develop item specifications	Incorporate feedback from the stakeholder review into the EPEs. Content and accessibility specialists finalize the EPEs and draft the item specifications, including prototype/sample items. Completed by Cognia development team	January–March 2017
Stakeholder review: draft item specifications and prototype/sample items	Convene stakeholder committee to review a draft set of item specifications and corresponding prototype/sample items at each of the grade levels. Gather feedback to refine the item design and item specifications. Review completed by MSAA Science Partners and Cognia content and accessibility specialists external to MSAA Science team	March 2017

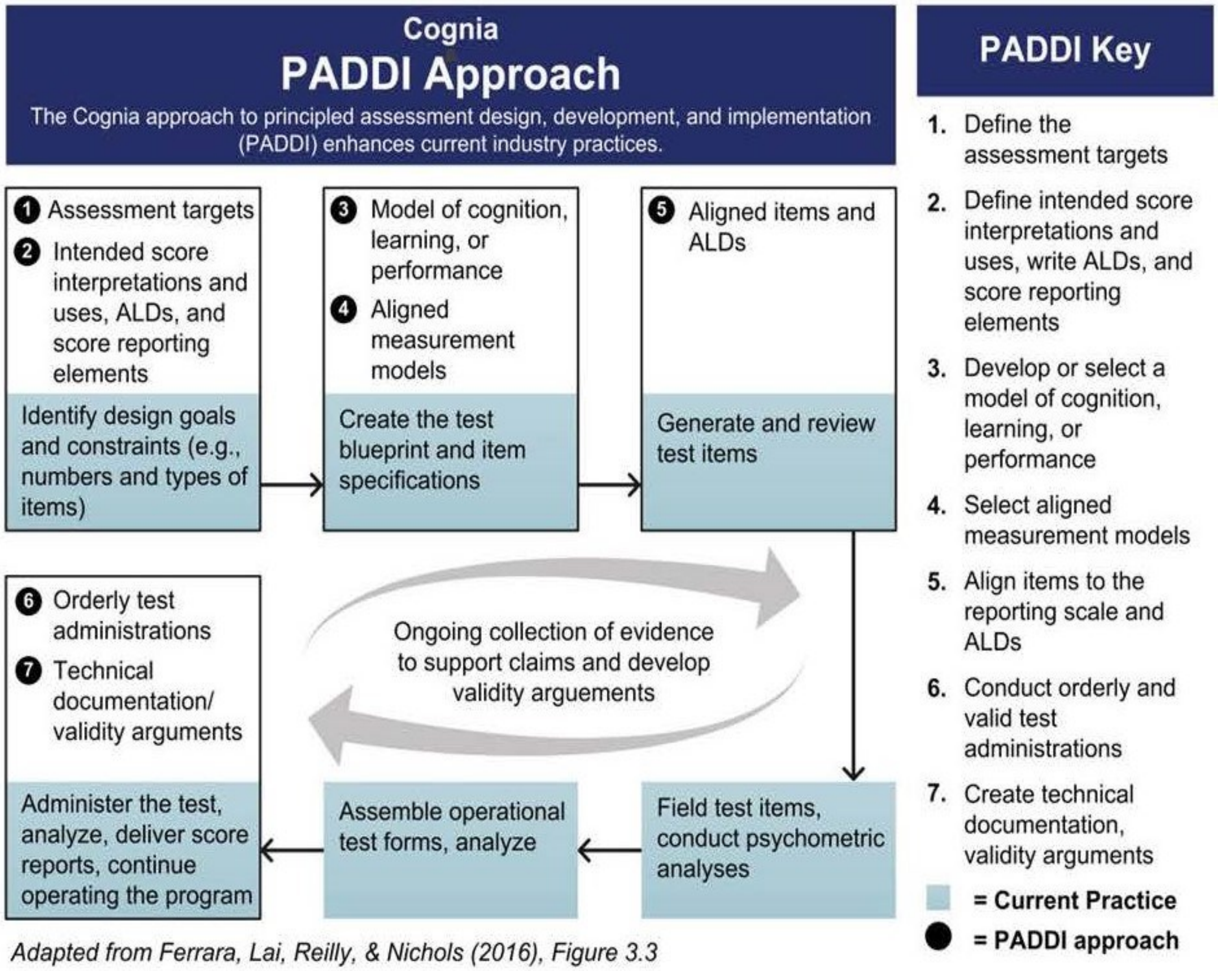
continued

Activity	Purpose & Stakeholders Involved	Timeframe*
Refine and finalize item specifications and develop items	Content and accessibility specialists complete multiple iterative rounds of refinement of the item specifications and begin item development. Following initial item development, review item specifications further and revise. Item development is ongoing. Completed by Cognia development team	Spring 2017 and ongoing
Stakeholder review: item content and bias-sensitivity meetings	Convene stakeholder committee to review items and provide content specific feedback as well as bias-sensitivity feedback. Occurs over multiple years as new item development happens. Reviews completed by MSAA Science Partners, MSAA Science state/entity educators, and Cognia content and accessibility specialists external to MSAA Science team	September 2018 March 2021 April 2022 April 2023
Field test administration	Field test administration with the MSAA Science Partners.	Spring 2021
Operational administration	Operational administration with the MSAA Science Partners.	Spring 2022

**Notably, the first operational administration of the MSAA Science was planned for the spring of 2020, but it did not occur due to school closings in response to COVID-19.*

The release of *A Framework for K-12 Science Education* in 2012 provided a national focus on moving toward multidimensional science instruction that fully integrated the Science and Engineering Practices (e.g., planning and carrying out investigations) and Crosscutting Concepts (e.g., patterns) with typical science content within the Disciplinary Core Ideas (e.g., forces and motion) rather than instruction of these concepts in discrete segments. Standards based on the *Framework*, such as the Next Generation Science Standards (NGSS, 2013) or NGSS-like state content standards, weave the practices, crosscutting concepts, and core ideas together through Performance Expectations (PEs). As states were developing general education assessments based on NGSS or NGSS-like standards, it was also important to develop similar alternate assessments that would allow students with significant cognitive disabilities to demonstrate their knowledge and skills of multidimensional science concepts. The Cognia development team followed a principled assessment design process and utilized the published national resources of the *Framework* and NGSS. As outlined in Ferrara, Lai, Reilly, and Nichols (2016), “principled approaches provide concepts, procedures, and tools to guide assessment design, development, and implementation decisions” (p. 3).

Figure 2-1. Cognia's Approach to Principled Assessment Design, Development, and Implementation



The first steps involved the Cognia development team developing a claim, assessment targets (e.g., test design, test blueprints), policy performance level descriptors (PLDs), and score reporting elements, followed by the selection of PEs and creation of the Extended Performance Expectations (EPEs). The Cognia development team looked to research in special education, specifically focusing on students with significant cognitive disabilities that has provided models to support how learning opportunities and assessment tasks can be designed to provide evidence for inferences about what students know and what they can do across a full range of performance (Kleinert, Browder, Towles-Reeves, 2009). The Cognia development team used this research along with experience developing other alternate assessments, such as the National Center and State Collaborative (NCSC) and its continuation in the Multi-State Alternate Assessment (MSAA) as helpful guidelines for developing the MSAA Science. Utilizing experts in both the science content area (specifically in-depth knowledge of the *Framework* and NGSS) and special education (specifically teaching students with significant cognitive disabilities), as well as psychometricians familiar with the nuances of alternate assessments, the Cognia development team developed the science alternate assessment design based on the assessment constructs and content-model definitions, which assume students can learn (1) when given the opportunity to learn multidimensional science academic standards in elementary, middle, and high school grades, and (2) when the prioritized assessment constructs focus on the critical content for progressing through the grade spans with each building on the subsequent grade span and using structured scaffolds and supports that do not interfere with the measurement of the science content.

The Cognia development team drafted the policy PLDs with four performance levels that describe expectations of student knowledge and skills at each level based on the assessment's claim. Based on the student population and design of the assessment, the Cognia development team and psychometricians determined reporting would be at the content-area level including scale score and performance-level designation (i.e., Level 1, Level 2, Level 3, Level 4), and the measurement model would utilize item response theory (IRT), specifically two-parameter logistic (2PL) model. Additional detailed information about reporting and IRT scaling and equating is available in Chapters 7 and 9 respectively. The following important elements guided development of the science policy PLDs and informed the selection of prioritized assessment content by the Cognia development team:

- a focus on the levels of complexity, the depth and breadth, and the accuracy of understanding needed at each performance level;
- the need for scaffolds and supports for students with significant cognitive disabilities to permit independent demonstration (without changing the content being assessed);
- incorporation of multidimensional science content (core ideas, practices, crosscutting concepts) that ranges from concrete to abstract concepts; and
- the level of support, interrelated with the science content, to help distinguish the performance levels. For example, a student performing at the highest level may only need minimal scaffolding on more abstract concepts being measured, whereas a student performing at one of the mid-levels may need additional support built into the item and focus on more concrete concepts, and a student performing at the lowest level may require the additional supports along with content assessed that is focused on only one dimension of the science concepts (e.g., core ideas).

According to the *Standards for Educational and Psychological Testing*, tests should be designed to minimize construct-irrelevant barriers for all test takers in the target population (AERA et al., 2014, pp. 6–7). The Cognia development team adhered to universal design principles and considerations related to accessibility and item features throughout the next phase of developing the MSAA Science. Universal Design for Learning (UDL) seeks to optimize the accessibility of educational materials and assessments while minimizing separate-but-equal situations. To allow the widest possible range of students to demonstrate what they know and can do, and to be able to make valid inferences about the performance of all students who participate in an assessment, universally designed assessments are developed from the beginning with an eye toward maximizing fairness (Johnstone, Altman, & Thurlow, 2006). The Cognia development team applied their understanding of the characteristics of this student population and UDL principles to inform the design of each item. The development team focused on minimizing the likelihood that any necessary additional adaptations and accommodations would interfere with the measured construct. A strength of the principled design approach that they followed was the support it provided for the development of items that (a) focused on construct-relevant content (the knowledge, skills, and abilities intended to be assessed), (b) minimized the impact of construct-irrelevant skills (e.g., inability to read text due to size of print, inability to access items due to absence of assistive device, inability to engage with the items), and (c) considered appropriate accessibility options (Cameto, Haertel, Morrison, & Russell, 2010, p. 1). Accessibility and assessment features of the MSAA online delivery system and structured administration guidelines were built from the foundational work of NCSC, which created and adopted policies specific to accessible and flexible assessment delivery (e.g., computer-based, paper-based, or hybrid administration model, zoom, masking, etc.). These features are also part of the MSAA Science as it is part of the MSAA comprehensive program. The Cognia development team built the accessibility and item features of the item design, such as depth of knowledge, text complexity, context, and degree and type of scaffolds and supports, into the item specifications to establish a defined

consistency within the item development efforts. The Cognia development team selected the PEs at each grade level based on the claim and assessment constructs that would cover a range of content knowledge in each of the science domains (i.e., Physical Sciences, Life Sciences, and Earth and Space Sciences) while being manageable for the student population and would yield reliable scores that demonstrate knowledge, skills, and understanding of the science concepts. The EPEs developed at three complexity levels serve as varied assess points to the emphasis in the PEs. A key tenet of the *Framework* is multidimensional expectations. As such, the EPEs also sought to uphold that tenet. The Level 3 EPE closely resembles the multidimensional elements of the PE, with Levels 2 and 1 EPEs also maintaining multidimensional elements, when possible, for the science concept and structured in a way that builds skills from Level 1 to Level 2 and ultimately to Level 3.
Following the principled approach to develop items that held true to the claim and content-model definitions, Cognia's development team created detailed item specifications for each domain and grade level. The secure item specifications include:

- the Performance Expectation text and corresponding Science and Engineering Practice, Disciplinary Core Idea, and Crosscutting Concept that comprise the Performance Expectation;
- the Extended Performance Expectations (three levels) and corresponding Science and Engineering Practice, Disciplinary Core Idea, and Crosscutting Concept that comprise the Extended Performance Expectations (target and supporting, as appropriate);
- target vocabulary (terms eligible and non-eligible to be assessed at each level as well as overall expected level of vocabulary coding of the EPE by level);
- additional clarifications and assessment boundaries/content limits (additional information to consider about focus and intent of the EPE by level, as well as any limitation as to what may be assessed);
- typical stimulus elements (elements/format of the information typically presented in a stimulus for a particular EPE by level);
- sample contexts (guidance for the overall familiarity and complexity of contexts for items at that EPE by level, accompanied by several sample contexts that would be appropriate for items for that EPE);
- target Depth of Knowledge (DOK) and descriptors (provides the DOK level generally expected to be targeted for the EPE by level);
- sample stems (one or more brief examples of items aligned to each EPE by level);
- response types (types of option formats that are likely to be presented for the items);
- distractor rule (describes how related the distractors should or should not be to the stimulus);
- scoring rules (describes the number of points a student may be awarded for an item).

The assessment items incorporate important aspects of item design related to both varying levels of content complexity and the types of scaffolds and supports. Additional detailed information about the PEs, test blueprints, EPEs, and items is available in Chapter 3.

The Cognia development team convened stakeholder reviews at various points along the development of the MSAA Science as shown in Figure 2-1 in order to gather feedback and validation of the work/development being done. According to AERA, APA, and NCME (2014), the review process should include expert judges to review items, qualifications, and relevant experiences; in addition, demographic characteristics, item review instructions, and reviewers' training should be documented (pp. 87–88). Cognia's development team collected evidence in support of these requirements. For each stakeholder review, the Cognia development team conducted a training and security protocol overview and provided specific materials including feedback focus questions customized to the particular stakeholder review. The Cognia development team gathered feedback and incorporated it into the various documents reviewed. A list of stakeholders (which included Cognia content and accessibility specialists external to the MSAA Science team, MSAA Science Partners, and MSAA Science state/entity educators) who participated in the reviews can be found in Appendix B.

The Cognia development team first conducted a stakeholder review of the assessment targets, specifically the test blueprints, selected PEs, and the alignment of the EPEs to the PEs and across EPE levels. Cognia content and accessibility specialists external to the MSAA Science team participated in this review. As noted in the Links for Academic Learning (LAL) model developed by Flowers, Wakeman, Browder, and Karvonen (2007), the alignment aspects addressed were as follows:

- Domain coverage: coverage by EPE levels for the range of prioritized PEs
- Content centrality: the focus of achievement in the EPE levels maintains fidelity with the content of the prioritized PEs
- Performance centrality: the focus of achievement in the EPE levels maintains fidelity with the specified performance of the prioritized PEs
- Depth of Knowledge (DOK): a range/progression reflected in the EPE levels and its ability to maintain fidelity to the DOK of the prioritized PE

Stakeholders first reviewed the test blueprint domain percentages and rationale for the selected PEs. Secondarily, stakeholders reviewed the alignment of the EPEs to the PEs, to confirm the selected PEs were appropriate for each grade and that there were no gaps in emphasis (and, if gaps were identified, to provide recommendations to adjust the prioritized PEs). Stakeholders then reviewed the EPEs for a clear progression in student understanding across the EPE levels and whether the differences in complexity between the levels were appropriate for a progression across the levels. Lastly, stakeholders considered the multidimensional nature of the science, to determine whether the EPEs appropriately targeted the core ideas, practices, and crosscutting concepts. If stakeholders determined there was no link between the EPEs and the PEs, if there were gaps in the PEs selected, or if the EPEs did not meet the desired progression and multidimensional designation, then specific suggested edits were requested to be provided by the stakeholders. Overall, the stakeholders confirmed the domain coverage, content centrality, and performance centrality for the blueprint. In their EPE review the stakeholders' provided suggestions related to the wording of the EPEs and the progression of content expectations in the EPE levels. Feedback on wording centered primarily on clarity and intentionality (e.g., is the expectation clearly stated, and is it eliciting the knowledge or skill intended versus inadvertently requiring something else less relevant or too difficult). Feedback on the progression of the EPE levels focused on the clarity and appropriateness of the progression from Level 1 to Level 2 to Level 3 in the EPE. The Cognia development team incorporated the feedback in all cases in which they determined the feedback suggestions to be appropriate in accordance with the intention of the assessment constructs and measurement.

The Cognia development team also conducted a stakeholder review of the detailed item specifications per grade level and the item prototypes. MSAA Science Partners and Cognia content and accessibility specialists external to MSAA Science team participated in this review. The reviewers focused on information within the item specifications, another alignment review of the EPEs similar to what was noted in the previous stakeholder review, and on the fidelity of the specifications on actual item prototypes. The review checklist used (Figure 2-2) was as follows:

Figure 2-2. Stakeholder Review Checklist

Evaluate the following for each item set:	
Alignment	Alignment to the NGSS standards known as Performance Expectations (PEs) • Do the items align to the PEs? • Do the items assess the intent of the **EPEs**? • Are the **Disciplinary Core Ideas (DCIs)** integrated in each EPE item as stated? • Are the **Science and Engineering Practices (SEPs)** integrated in each EPE item as stated? • Are the **Crosscutting Concepts (CCCs)** integrated in each EPE item as stated? • Do the items align to the **Item Specifications**?
Item Content	Content • Is the content accurate? • Is the content consistent with NGSS intended instruction? Teacher Script • Is the language consistent? • Are the directions for the teacher clear? • Is the “Prepare PBT” and “Prepare CBT” text consistent and clear? • Are the “SAY” and “ASK” portions that are spoken to the student clear? • If there is no “SAY” text, should there be? Student Response • Is there only one correct answer listed? Graphics • Do the graphics support the content? • Are the graphics simple and free of distractions?
Complexity	Depth of Knowledge (DOK) • Do you agree with the assigned rating for DOK? Presentation Rubric • Do you agree with the assigned ratings for each component of the Presentation Rubric (Volume of Information, Vocabulary, and Context)? Progression • Is there a clear progression from Level 1 items to Level 2 items to Level 3 items? • Is there a clear progression from one grade span to the next?
Bias & Sensitivity	Items should avoid Economic, Regional, Cultural, or Gender Sensitivity. • Is the topic/content too specific to one particular state or region? • How might the topic/context affect a student who recently had a personal experience with the subject (e.g., flooding, hurricanes, or other weather events)? • Is the topic/context grade appropriate?

Each reviewer provided feedback via a feedback form. In addition, the Cognia development team held a virtual meeting with the stakeholders so that reviewers each had an opportunity to discuss their feedback with the Cognia development team and the other reviewers. A high-level view of the feedback received focused on the following:

- clarifying the requirements of the PE and EPE;
- confirming or revising the Depth of Knowledge (DOK) designation;
- identifying any assessment boundaries;
- item progression across EPE Levels 1, 2, and 3;
- alignment;
- item clarity;
- simplifying the item when possible;
- vocabulary accessibility;
- language load for this population;
- graphics;
- confirming the correct response option is the only correct response option;
- when and where to provide graphic descriptions within the item;
- clarifications to add to the Editorial and Graphic Style Guides;
- alternate text considerations; and
- identifying any bias and/or sensitivity concerns.

The Cognia development team analyzed this item specifications feedback and item-specific feedback. The development team then incorporated edits in the item specifications and updated the item prototypes.

Once the item specifications were finalized following the stakeholder review, the Cognia development team began the work on developing the assessment items. Cognia held Item Content and Bias-Sensitivity reviews for each development cycle in which new items were created. At the start of each development cycle, the Cognia development team reviewed the item specifications and updated them as needed based on lessons learned from previous stakeholder reviews and field-testing results. For each Item Content and Bias-Sensitivity review, MSAA Science Partners, MSAA Science state/entity educators, and Cognia content and accessibility specialists external to MSAA Science team participated. Cognia organized the reviewers into grade-specific panels and provided the following materials:

- Grade-level item specifications, which include the PEs and EPEs
- Review checklist
- DOK chart
- PDFs of items

Facilitators from the Cognia development team asked reviewers to evaluate the alignment, item content, complexity, and bias and sensitivity considerations for each item reviewed. The Item Content and Bias-Sensitivity Review Checklist (Figure 2-3) was as follows:

Figure 2-3. Item Content and Bias-Sensitivity Review Checklist

Evaluate the following for each item set:	
Alignment	Alignment to the PEs • Do the items align to the PEs? • Do the items assess the intent of the **EPEs**? • Are the **Disciplinary Core Ideas (DCIs)** integrated in each EPE item as stated? • Are the **Science and Engineering Practices (SEPs)** integrated in each EPE item as stated? • Are the **Crosscutting Concepts (CCCs)** integrated in each EPE item as stated? • IF used, are Supporting Science and Engineering Practices (SEPs) and Crosscutting Concepts (CCCs) integrated in each EPE item as stated?
Item Content	Content • Is the content accurate? • Is the item scenario accessible, clear, and relevant to the student population that includes a wide range of disabilities? • Is the item text clear and succinct, avoiding words with multiple meanings and extraneous language? Student Response • Is there only one correct answer listed? • Are the response options clear, appropriate, and comparable in length? Graphics • Do the graphics support the content? • Are the graphics simple and free of distractions? • Does the alternative text (**boldfaced** in Item Table) provide contextual information that makes graphics accessible to students with visual impairments/blindness and the most significant cognitive disabilities? • Does the alternative text clearly align with what is represented in the corresponding graphic? Item Table • Are the "**SAY**" and "**ASK**" portions that are spoken (**boldfaced**) to the student clear and match the text in the stimulus? • Are the "**SAY**" and "**ASK**" portions that provide teacher directions (*italicized*) clear?
Complexity	Depth of Knowledge (DOK) • Is the assigned DOK rating appropriately represented? Progression • Is there a clear progression from Level 1 items to Level 2 items to Level 3 items?
Bias & Sensitivity	Items should avoid Economic, Regional, Cultural, or Gender Sensitivity. • Does the item content present an unfair advantage or disadvantage to any group of students? • Will the topic have an unintended impact on the student who recently had a personal experience with the subject (e.g., weather-related event, etc.)? • Is the content appropriate in terms of familiarity, interest, age, and grade? • Does the item avoid content that is offensive to any group (based on race, gender, sexual orientation, age, religion, ethnicity, socioeconomic status, or regional origin)?

The Cognia Item Content and Bias-Sensitivity facilitator recorded reviewer feedback using PDF mark-ups for each item. Consensus agreement was not necessary among all reviewers, but the facilitator sought overall agreement with feedback and recorded dissenting or differing perspectives in the PDF. Following the Item Content and Bias-Sensitivity review, Cognia's development team reconciled the reviewer feedback and incorporated it into the items to prepare them for field testing.

Additional information about current stakeholder involvement and stakeholder reviews is available in the next section and Chapter 4 respectively.

2.3 Current Stakeholder Involvement

Several stakeholders are involved in the continuing development of the MSAA Science. American Samoa, Arizona, Bureau of Indian Education (BIE), Commonwealth of the Northern Mariana Islands (CNMI), Guam, Maine, U.S. Virgin Islands, and Vermont collaborate with Cognia on the MSAA Science. Members of this body provide input and feedback on specific aspects of the assessment. Certain activities are specific to MSAA Science, while others apply to the comprehensive MSAA program.

The MSAA Science Partners' involvement includes participating in development planning, item reviews, post-item review committee reconciliation, and data review summaries. Input is also provided for the overall administration during test-construction form planning and reviews of the computer-based and paper-based administration materials. Specific feedback is also provided as it relates to the MSAA Science reports and their design. In addition to the MSAA Science Partners, stakeholders from schools and districts across the MSAA Science Partners participate in the field-test item development process during the Item Content and Bias-Sensitivity meeting. Additional detailed information is available in Chapter 4.

The MSAA Science Partners are part of the full collaborative of MSAA Partners. The MSAA Partners structure decision-making authority into various subcommittees that MSAA Partners elect to be a part of. Overall, MSAA Partners oversee development of the Test Administration Manual (TAM), MSAA Online Assessment System User Guide for Test Administrators, MSAA Online Assessment System User Guide for Test Coordinators, Parent Guides in English and Spanish, online training modules, and final quiz for TAs. Additionally, they are responsible for decisions on the overall layout of the student results files and the final processing and reporting business requirements. They oversee planning Technical Advisory Committee (TAC) meetings, contribute to psychometric decisions, provide the content of the End-of-Test Survey, determine relevant policies, receive the survey results after administration, and make recommendations on the structure of the technical report. The MSAA Partners have determined development priorities for the online assessment platform used in the 2023 comprehensive MSAA program on an as-needed basis. This group has also reviewed recommendations and development pertaining to the security of the online platform and ultimately approved all changes made to the platform.

2.4 MSAA Participation

The criteria for student participation in the 2023 MSAA Science reflects the pervasive nature of a significant cognitive disability. All content areas are considered when determining who should participate in this assessment. Table 2-2 shows the participation criteria and the descriptors used to determine eligibility for each student. Appendix C shows the 2023 summary of tested students by demographic category.

Table 2-2. Participation Criteria

Participation Criteria	Participation Criteria Descriptors
The student has a significant cognitive disability.	Review of student records indicates a disability or multiple disabilities that significantly impact intellectual functioning and adaptive behavior. *
The student is learning content linked to grade-level content standards.	Goals and instruction listed in the individualized education program (IEP) for the student are linked to the enrolled grade-level content standards and address knowledge and skills that are appropriate and challenging for the student.
The student requires extensive, direct, individualized instruction and substantial supports to achieve measurable gains in the grade- and age-appropriate curriculum.	The student (a) requires extensive, repeated, individualized instruction and support that is not of a temporary or transient nature, and (b) uses substantially adapted materials and individualized methods of accessing information in alternative ways to acquire, maintain, generalize, demonstrate, and transfer skills across multiple settings.

**Adaptive behavior is defined as essential for someone to live independently and to function safely in daily life.*

Assessments for students with the most significant cognitive disabilities rely on a foundation of communicative competence. Students who have not yet developed consistent receptive and expressive communication are unlikely to be able to demonstrate what they know and can do on an assessment. Students who do not have a consistent appropriate mode of communication are identified during the assessment process. In order to meaningfully participate in the MSAA, students must be able to demonstrate communicative competence through an observable response mode. An observable response mode is a predictable and consistent behavior or movement that can be understood by a communication partner as intentional communication. The Student Response Check (SRC) is a task during which a student is asked to demonstrate their preferred mode(s) of communication. In these cases, the SRC aids in gathering information that is needed to determine whether there are communication barriers to meaningful participation in the MSAA assessment. If a student's responses to test items are not clearly observable or understood by the TA or scribe, the testing experience may need to be ended early. This process is called the Early Stopping Rule (ESR). In order to end the test for a student, the ESR procedures must be followed. For additional information on how the ESR data are provided to the MSAA Science Partners, districts, schools, and parents/guardians, see Chapter 7.

Figure 2-4 shows the procedure for determining if the SRC is appropriate to administer and, if so, how to proceed in determining if the student has an observable, interpretable mode of communication to use throughout testing. If clear, intentional communication is not shown, the ESR may be applied. This figure shows the process of implementing the ESR.

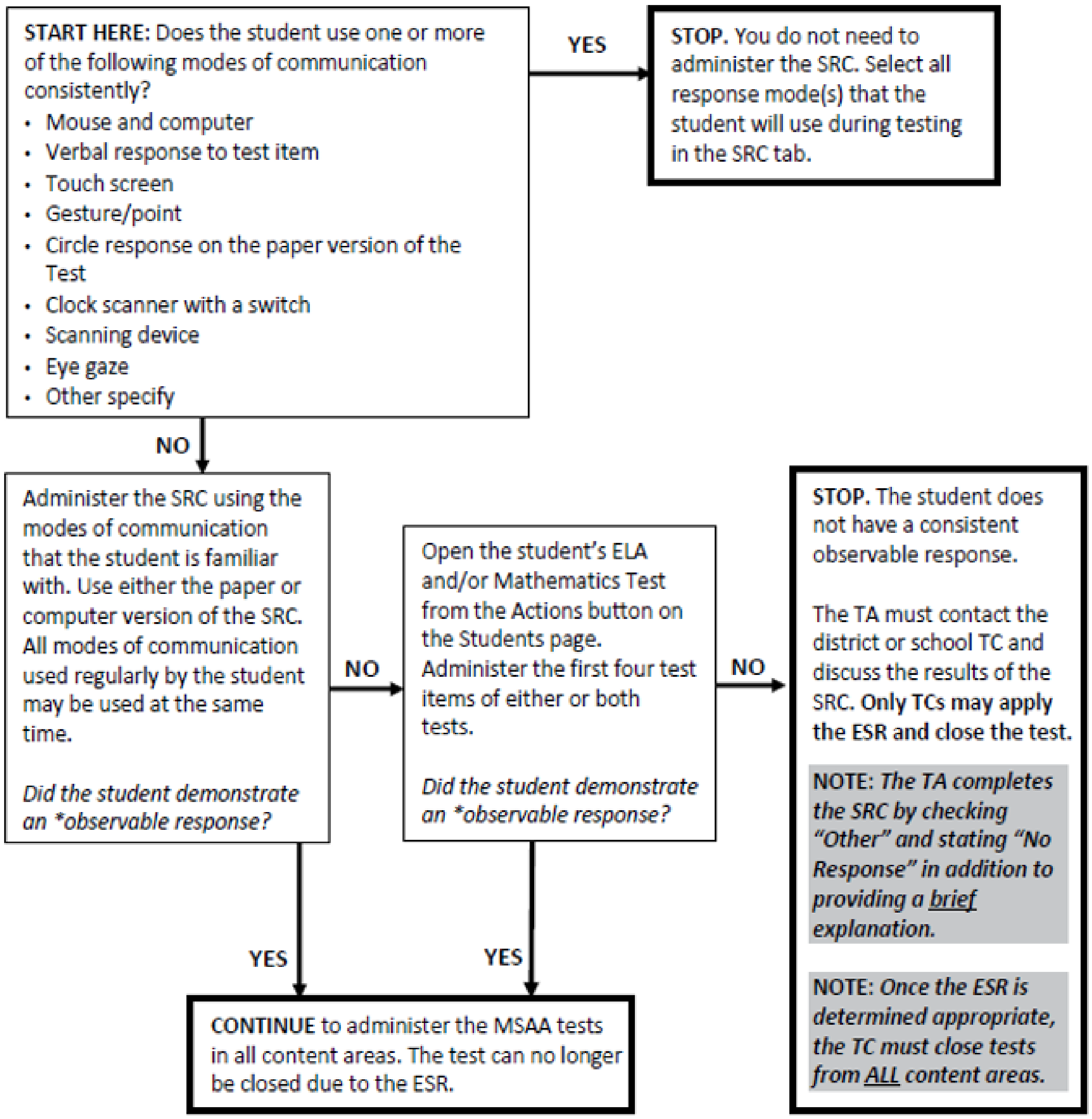


Figure 2-4. Student Response Check (SRC) Flowchart: When to Apply the ESR

The MSAA Science Partners provide very specific training to TAs on understanding the SRC and applying the ESR. Additionally, professional development is provided to TAs through a best practice module specific to the SRC and ESR. Detailed information regarding this is available in Chapter 5.

As an additional resource, teachers can use the Communication Tool Kit developed by NCSC to help these students develop an appropriate mode of communication. The Tool Kit consists of a series of professional development modules addressing seven component parts: identifying communication, considering sensory and motor factors, selecting communication targets, selecting AAC, teaching communication targets, embedding communication into academics, and monitoring performance. The Tool Kit can be found here: https://wiki.ncscpartners.org/index.php/Communication_Tool_Kit.

Chapter 3. Test Development: Content and Administration

3.1 History of Three-Dimensional Science Standards and Extended Performance Expectations (EPEs)

The MSAA Science aligns to and assesses academic standards for three-dimensional science standards that are appropriate for the student population. Three-dimensional science standards, such as the Next Generation Science Standards (NGSS), are based on *A Framework for K-12 Science Education* (National Research Council, 2012). Standards based on the *Framework* are complex science standards (often termed Performance Expectations, or PEs) that integrate three dimensions in each standard: Disciplinary Core Ideas (DCIs), Science and Engineering Practices (SEPs), and Crosscutting Concepts (CCCs).

The DCIs included in the *Framework* represent the science content ideas from Physical Sciences, Life Sciences, Earth and Space Sciences, and Engineering Design that are considered to be most central to science education. The writers of the *Framework* chose the DCIs for being major organizing principles of the disciplines, providing key tools for understanding or investigating more complex ideas and solving problems, and relating to the interests, personal and/or societal concerns, and life experiences of students (*Framework*, p. 31). While the *Framework* focuses on a more limited set of core ideas than past science standards did, resulting standards developed from the *Framework* still have a very large number of PEs in each grade span. To develop an appropriate science assessment for students with the most significant cognitive disabilities (SCD), the Cognia development team (composed of content specialists and accessibility specialists) had to further limit the number of standards being assessed.

The Cognia development team chose a total of 12 PEs (based on the *Framework* and the NGSS) as the focus for each grade test. Another science alternate-assessment offering in the market assesses 9-15 science standards depending on the grade. The total standards per grade is also comparable to the number of standards assessed per grade in ELA and mathematics in the MSAA assessments.

In identifying these PEs for the MSAA science assessment, the development team's goal was to provide reasonable representation across the DCIs and to focus on the most fundamental, broad principles that would be accessible and meaningful for this student population as a progression from elementary to middle to high school. In support of that, the selected PEs represent content from all grades in the grade band for each test. For the grade 5 test, for example, PEs from grades 3, 4, and 5 are included; the progression of standards in those grades is such that to provide a solid representation of the core ideas and understandings that students need to progress from elementary school to middle school, the PEs needed to be selected across grades. Likewise for the grade 8 and high school test (even though the standards are presented as grade band in these levels), the selected PEs would typically be taught across multiple years in middle school and high school, respectively. The development team, composed of science content specialists and accessibility specialists, used research in special education and student learning, their experience developing other alternate assessments in science, and their direct prior experience with students in the classroom to support their final PE decisions. The development team also compared their selections to those of another emerging *Framework*-based science alternate assessment. Further information regarding the history of the MSAA Science test design including stakeholder reviews is available in Chapter 2. The following Table 3-1 shows the collection of PEs chosen to be assessed on the grade 5 test.

Table 3-1. Selected Performance Expectations for Grade 5 Test

Performance Expectation (PE)	DCI	SEP	CCC
5-PS1-2. Measure and graph quantities to provide evidence that regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is conserved.	PS1.A–Structure and Properties of Matter PS1.B–Chemical Reactions	Using Mathematics and Computational Thinking	Scale, Proportion, and Quantity
3-PS2-2. Make observations and/or measurements of an object's motion to provide evidence that a pattern can be used to predict future motion.	PS2.A–Forces and Motion	Planning and Carrying Out Investigations	Patterns
5-PS2-1. Support an argument that the gravitational force exerted by Earth on objects is directed down.	PS2.B–Types of Interactions	Engaging in Argument from Evidence	Cause and Effect
4-PS3-4. Apply scientific ideas to design, test, and refine a device that converts energy from one form to another. *	PS3.B–Conservation of Energy and Energy Transfer PS3.D–Energy in Chemical Processes and Everyday Life ETS1.A–Defining and Delimiting an Engineering Problem	Designing Solutions	Energy and Matter
5-PS3-1. Use models to describe that energy in animals' food (used for body repair, growth, and motion, and to maintain body warmth) was once energy from the sun.[1]	PS3.D–Energy in Chemical Processes and Everyday Life	Developing and Using Models	Energy and Matter
4-LS1-1. Construct an argument that plants, and animals have internal and external structures that function to support survival, growth, behavior, and reproduction.	LS1.A–Structure and Function	Engaging in Argument from Evidence	Systems and System Models
3-LS3-1. Analyze and interpret data to provide evidence that plants and animals have traits inherited from parents and that variation of these traits exists in a group of similar organisms.	LS3.A–Inheritance of Traits LS3.B–Variation of Traits	Analyzing and Interpreting Data	Patterns
3-LS4-1. Analyze and interpret data from fossils to provide evidence of the organisms and environments in which they lived long ago.	LS4.A–Evidence of Common Ancestry and Diversity	Analyzing and Interpreting Data	Scale, Proportion, and Quantity
5-ESS1-2. Represent data in graphical displays to reveal patterns of daily changes in the length and direction of shadows, day and night, and the seasonal appearance of some stars in the night sky.	ESS1.B–Earth and the Solar System	Analyzing and Interpreting Data	Patterns
3-ESS2-1. Represent data in tables and graphical displays to describe typical weather conditions expected during a particular season.	ESS2.D–Weather and Climate	Analyzing and Interpreting Data	Patterns
5-ESS2-1. Develop a model using an example to describe ways in which the geosphere, biosphere, hydrosphere, and/or atmosphere interact.	ESS2.A–Earth Materials and Systems	Developing and Using Models	Systems and System Models
5-ESS3-1. Obtain and combine information about ways individual communities use science ideas to protect the Earth's resources and environment.	ESS3.C–Human Impacts on Earth Systems	Obtaining, Evaluating, and Communicating Information	Systems and System Models

**PEs marked with an asterisk incorporate Engineering constructs.*
[1]This PE crosses Physical Sciences and Life Sciences. It will be classified in Life Sciences in fulfilling the blueprint distribution.

For the grade 5 test, the Cognia development team selected PEs from grades 3–5 to generate the best representation of broad, fundamental principles for the elementary grade assessment. Because the NGSS spread science topics out across grades in the 3–5 grade band, there are some key content ideas for forces and motion, as well as for heredity, biological evolution, and weather, that only appear in standards at grade 3. Although the test is targeted for administration to students in grade 5, the development team included the following four grade 3 PEs to ensure focus on all foundational areas that students would need exposure to, to prepare for middle school expectations:

- 3-PS2-2 focuses on basic patterns of motion, as a foundation of the cause-and-effect exploration of forces and motion. This PE also provides an opportunity to expose students to the various types of forces, from physical contact forces to gravity and magnetism, linking to another motion/forces PE within elementary and to other motion/forces PEs in later grades.
- 3-LS3-1 introduces the fundamental principle of inheritance of traits (traits passed from parents to offspring) as well as the idea of variation, which are both cornerstones of the study of genetics and biological evolution.
- 3-LS4-1 provides an accessible foundation for thinking about evidence of organisms' fit to the environment, and changes in organisms and environments over time.
- 3-ESS2-1 focuses on the most foundational understandings of weather, which are then extended in other elementary PEs and in later grades in studying interactions of Earth's systems, geoscience processes changing Earth's surface, water cycling through Earth's systems, and the larger concept of climate.

Note that while the chosen PEs may seem to lean more toward Physical Sciences than Life Sciences (5 PEs coded to Physical Sciences and only 3 PEs coded to Life Sciences), PE 5-PS3-1 is a "crossover" PE that connects the physical science concept of energy in everyday life with the life science concept of matter and energy flow. Although 5-PS3-1 has a physical science coding, educators would typically teach the standard within an ecology unit (and is therefore classified as a Life Science PE in the test blueprint).

Additionally, there are no PEs in the elementary grade test for Physical Sciences DCI PS4, Waves and Their Application in Technologies for Information Transfer. The concept of waves is abstract, and thus, based on student learning patterns in this population, the development team viewed the core idea as more appropriate to address in the grade 8 test than in this grade band for students with significant cognitive disabilities. Likewise in Life Sciences, although no performance expectation is explicitly aligned to DCI LS2 (Ecosystems: Interactions, Energy, and Dynamics), PE 5-PS3-1 overlaps heavily with these concepts. All other DCIs are represented in the elementary grade test. Additional detailed information regarding the rationale of PE selections completed by the Cognia development team for all grades is provided in Appendix D.

As the Cognia development team selected and finalized the PEs for each grade band, they checked the progression of DCIs across grades to help validate the appropriateness of the collection of PEs chosen for assessment on each grade's test. The following tables show an example of the final prioritized PEs and associated DCIs for the Physical Sciences across grades 5, 8, and HS.

Table 3-2. Selected Performance Expectations for Physical Sciences Across Grades 5, 8, and HS

Grade	Performance Expectation (PE)	DCI
5	5-PS1-2. Measure and graph quantities to provide evidence that regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is conserved.	PS1.A PS1.B
	3-PS2-2. Make observations and/or measurements of an object’s motion to provide evidence that a pattern can be used to predict future motion.	PS2.A
	5-PS2-1. Support an argument that the gravitational force exerted by Earth on objects is directed down.	PS2.B
	4-PS3-4. Apply scientific ideas to design, test, and refine a device that converts energy from one form to another. *	PS3.B PS3.D ETS1.A
8	MS-PS1-2. Analyze and interpret data on the properties of substances before and after the substances interact to determine if a chemical reaction has occurred.	PS1.A PS1.B
	MS-PS2-2. Plan an investigation to provide evidence that the change in an object’s motion depends on the sum of the forces on the object and the mass of the object.	PS2.A
	MS-PS3-5. Construct, use, and present arguments to support the claim that when the kinetic energy of an object changes, energy is transferred to or from the object.	PS3.B
	MS-PS4-2. Develop and use a model to describe that waves are reflected, absorbed, or transmitted through various materials.	PS4.A PS4.B
HS	HS-PS1-2. Construct and revise an explanation for the outcome of a simple chemical reaction based on the outermost electron states of atoms, trends in the periodic table, and knowledge of the patterns of chemical properties.	PS1.A PS1.B
	HS-PS2-3. Apply scientific and engineering ideas to design, evaluate, and refine a device that minimizes the force on a macroscopic object during a collision. *	PS2.A ETS1.A ETS1.C
	HS-PS2-5. Plan and conduct an investigation to provide evidence that an electrical current can produce a magnetic field and that a changing magnetic field can produce an electrical current.	PS2.B PS3.A
	HS-PS3-2. Develop and use models to illustrate that energy at the macroscopic scale can be accounted for as a combination of energy associated with the motions of particles (objects) and energy associated with the relative positions of particles (objects).	PS3.A

**PEs marked with an asterisk incorporate Engineering constructs.*

Table 3-3. DCI Coverage for Physical Sciences Across Grades 5, 8, and HS

	Physical Sciences Disciplinary Core Idea (DCI) Coverage Across Grades												
	PS1.A	PS1.B	PS1.C	PS2.A	PS2.B	PS2.C	PS3.A	PS3.B	PS3.C	PS3.D	PS4.A	PS4.B	PS4.C
Grade 5	X	X		X	X			X		X			
Grade 8	X	X		X				X			X	X	
Grade HS	X	X		X	X		X						

Ultimately, the selected PEs within and across grades represent a content progression supporting essential learning and understandings in the Physical Sciences. The constructs of structure and properties of matter, chemical reactions, forces and motion, types of interactions, and conservation and transfer of energy are all well-represented across the grade bands. Additionally, basic understanding of waves and their behavior is included in grade 8.

Additional detailed information regarding the final prioritized PEs and associated DCIs for the Physical Sciences, Life Sciences, and Earth and Space Sciences across all three grades is provided in Appendix D.

The Cognia development team also examined the representation of SEPs and CCCs across the selected PEs. The development team wanted to ensure most, if not all, SEPs and CCCs were included for assessment in each grade test because of the importance the Framework places on the use of multiple practices and crosscutting concepts in instruction of the standards. As can be seen in Table 3-1 for grade, the selected PEs incorporate seven of the eight SEPs and five of the seven CCCs (with the other two CCCs not actually included in the elementary grade band in NGSS). For grade 8 the selected PEs incorporate five of the eight SEPs and all seven CCCs. For high school the selected PEs incorporate seven of the eight SEPs and six of the seven CCCs. While the majority of SEPs and CCCs were represented across the selected PEs in each grade, as described in section 3.1.1 the development team attended to opportunities to incorporate the remaining SEPs or CCCs into level 1 and level 2 access points where appropriate to an EPE progression.

Likewise, the Cognia development team included one or two engineering-aligned PEs in the selected PEs for assessment in each grade test. The *Framework* elevates the importance of engineering as equal to the traditional domains of physical, life, and earth and space sciences, and includes engineering constructs as DCIs in addition to SEPs. As can be seen in Table 3-1 for the grade 5 test, the development team included one engineering-aligned PE, 4-PS3-4. Grade 8 likewise includes one engineering-aligned PE, and high school includes two engineering-aligned PEs.

Prior to the start of item development, the Cognia development team created Extended Performance Expectations (EPEs). The EPEs are the standards developed to define academic grade-level content that is clearly linked to *Framework*-defined grade-level content, but at reduced complexity, breadth, and depth appropriate for students with significant cognitive disabilities. These EPEs represent the assessable grade-level content for the MSAA Science.

Section 3.1.1 provides additional details about the EPE development. Once the Cognia development team had drafted initial EPEs, the team convened a stakeholder group to review the test blueprints, test design, and draft version of the EPEs. Cognia content and accessibility specialists external to the MSAA team, with expertise in science and the assessment of students with significant cognitive disabilities, completed this review to evaluate the assessment design and collection of standards being assessed as appropriate and valid in support of the proposed assessment claim for the MSAA Science. MSAA Science Partners participated in a later review of item specifications and item prototypes built upon the EPEs and test design, thus allowing the MSAA Science Partners to evaluate a more operationally-defined design of the assessment.

3.1.1 Extended Performance Expectations (EPEs)

The Cognia development team extended each selected general education PE into three access points to create the EPEs. The operational items vary in complexity following those three access levels of the EPEs. The highest access point, Level 3, is intended to closely mirror the general education PE and represents the most cognitively demanding target for this student population. The Level 3 EPE is three-dimensional and aligned to the same three dimensions (DCI, SEP, and CCC) as the general education PE. However, the Level 3 EPE typically has a slightly lower cognitive demand than the general education PE. For example, the EPE may limit the number or types of examples that students will be expected to connect to the construct. Depending on the particular EPE and the phenomenon or context chosen for assessing the EPE, some items may not encompass all parts of the EPE, particularly if it has multiple examples/contexts listed. All Level 3 items will, however, align to all three dimensions and the essence of the EPE.

The other two access points, Level 2 and Level 1, are written as progression points that students would likely move through as they build proficiency towards the Level 3 EPE. Level 2 and Level 1 EPEs are therefore intended to provide a scaffold for instruction and learning by supporting students in the attainment of the target knowledge and skills expressed in the Level 3 EPE. All Level 2 EPEs are two-dimensional, occasionally three-dimensional. All Level 1 EPEs are one-dimensional (DCI), occasionally two-dimensional. This allows instruction and learning to be appropriately focused and scaffolded in a logical, accessible sequence towards the complex expectation of integrating multiple facets of science (DCI, SEP, and CCC) in the target performance.

Additionally, in evaluating alignment to the DCI in Level 1 items, and some Level 2 items, alignment is achieved through the use of the vocabulary and examples/contexts matching the DCI. For this population of students, simply being able to process the words and context of the DCI to answer the item is evidence of engagement with and use of DCI knowledge. Regarding the SEP and CCC, the Level 2 (and Level 1, if two-dimensional) access point may align to an SEP and/or CCC other than the ones in the general education PE and Level 3 EPE. The rationale for this is two-fold: first, this approach mirrors best practices for classroom instruction on three-dimensional science and supports an emphasis on including all SEPs and CCCs, not just select ones, over the course of instruction; and second, this approach allows a scaffolded progression towards proficiency to be developed for each EPE, rather than trying to differentiate nuances of the *degree* to which a student is demonstrating a singular target learning outcome. For an example of the EPEs, along with an explanation of the major layout components, see Appendix D.

3.2 Alignment and Linkages

Evidence that the test content reflects the concepts that were meant to be measured is one of the critical sources of information necessary to support valid interpretations of test scores (AERA et al., 2014). *Alignment* refers to coherent connections within and across a system (Forte, 2013a, 2013b). Traditional alignment procedures describe the degree of intersection, overlap, or relationship among academic content embedded in state content standards, assessment, and instruction (Webb, 2005).

As part of the assessment development process, ACS Ventures, LLC (ACS) conducted an external independent study to evaluate the alignment of the MSAA Science for grades 5, 8, and HS in fall 2022. ACS staff designed the study protocol based on information that Cognia provided about the MSAA Science assessment. ACS staff then facilitated the panelist meetings, analyzed the data from the meetings, and generated the alignment study report for Cognia and MSAA Science Partner reviews.

ACS used the Links for Academic Learning (LAL) alignment method developed by the National Alternate Assessment Center as the basis to conduct the content alignment review and analyze the results (Flowers et al., 2007). ACS adapted this method to best fit the MSAA Science data analysis needs. In addition to the LAL alignment questions, ACS also conducted supplementary analysis to confirm the EPEs to the designated dimensions (i.e., DCI, SEP, CCC). Overall, the study was designed to answer several key questions related to the alignment of the assessment. The alignment questions were:

1. What degree of *content centrality* is maintained between the items and the EPEs? (LAL Criterion 3)

2. What degree of *performance centrality* is maintained between the items and EPEs? (LAL Criterion 3)
3. Are there an adequate number of items representing each domain on the science alternate assessment test form (i.e., domain concurrence)? (LAL Criterion 4)
4. Does the collection of the science alternate assessment items represent multiple EPEs within each domain of the blueprints (i.e., range of knowledge)? (LAL Criterion 4)
5. Does the balance of representation indicate similar emphasis of science alternate assessment items as the blueprint? (LAL Criterion 4)
6. Is there a range of cognitive levels across the content of the science alternate assessment items and do the levels reflect fidelity with the cognitive levels in the EPEs? (LAL Criterion 4)
7. Is there a change in emphasis of age-appropriate content across grade levels (i.e., differentiation)? (LAL Criterion 5)
8. Is the content accessible to students with varying levels of communicative competence? (LAL Criterion 7)

The LAL method is appropriate for alignment of the EPEs to the corresponding general education PEs. The supplementary analysis is appropriate given the MSAA Science is intended to address multidimensional science concepts that include DCI, SEP, and CCC.

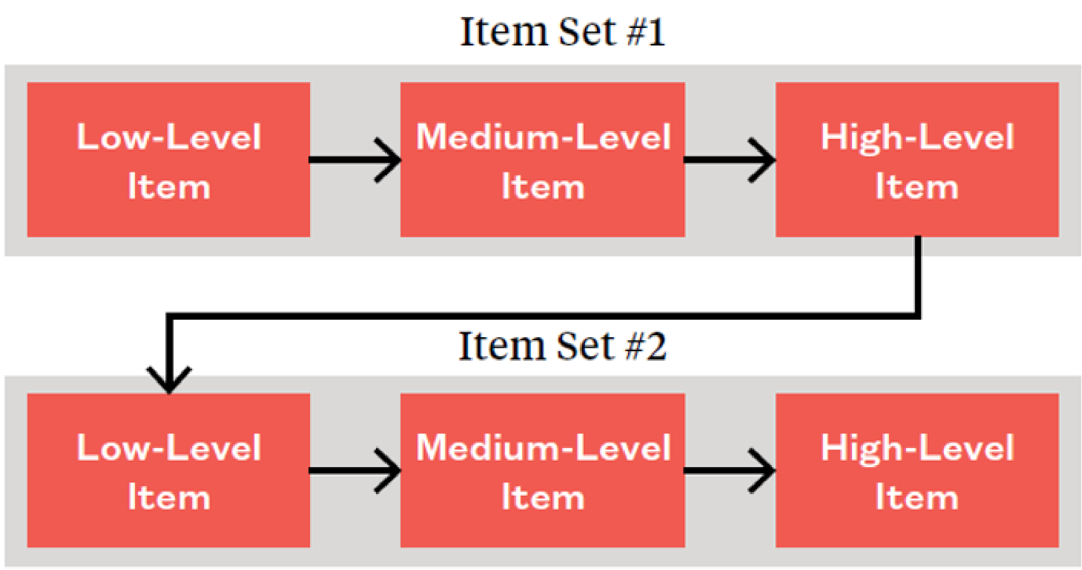
The MSAA Science items were designed to assess the knowledge and skills of a wide variety of students with the most significant cognitive disabilities, and the findings of the alignment study supported this. The grade-level items were rated as well-aligned for both content centrality and performance centrality, meaning at least 90% of items were judged as having some or all the same performance expectations of the EPEs. Domain concurrence for each grade level was rated as well-aligned, meaning at least 90% of the items on the test form align to an EPE defined in the blueprint, no item on the test form reflects expectations not defined in the grade level, and each of the domains in the blueprint is represented by items on the form. The range of knowledge and balance of representation for each grade level was rated as well-aligned. The reviewed items compared to the blueprint, EPEs, and point emphasis were varied for each domain represented (i.e., Physical Sciences, Life Sciences, and Earth and Space Sciences). The study also confirmed there is a range of cognitive complexity represented across the three levels in each grade. Differentiation across grade levels was rated as well-aligned for all grade levels, indicating an increased breadth and depth of the content. The alignment study results provided evidence that the assessment items allow access for students consistently using various communication modes and with specific characteristics. Panelists indicated that the accommodations and allowable modifications as defined in the TAM were accessible or somewhat accessible. For students who do not yet have clear, intentional communication even at the non-symbolic level, the assessment doesn’t have a way to capture responses. These students go through the SRC as defined in the TAM and likely have the ESR applied as an appropriate assessment procedure. Panelists indicated that the TAM clearly defined the accommodations, modifications, and supports allowed for standardized administration procedures. Additionally, the panelists confirmed strong alignment to the dimensions (i.e., DCI, SEP, CCC), providing evidence that the science content meets the multidimensional concepts as intended. The alignment study report is available at www.msaastates.com.

3.3 2023 MSAA Science Assessment Design

3.3.1 Operational Design and Administration

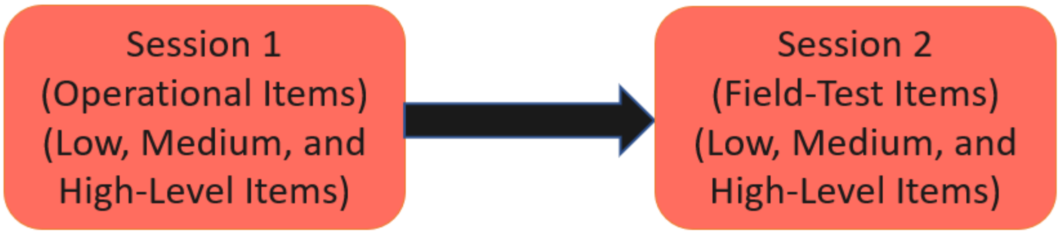
The operational MSAA Science is designed to produce scores that have a high degree of reliability and validity for the intended interpretation and uses. MSAA Science is composed of selected-response items, each with three response options. The assessment is developed and administered in item sets over two sessions. Item sets consist of three items, aligned to the three access levels of the Extended Performance Expectations (EPEs). The low-level item (Level 1) is administered first, the medium-level item (Level 2) is administered second, and the high-level item (Level 3) is administered third, and then a new set of items begins (as shown in Figure 3-1). There are also a small number of clusters, which contain two item sets (of three items each) bundled together with a single shared stimulus, to create a more integrated, performance task-like experience for the students.

Figure 3-1. Linear Administration of Item Sets



For the 2023 assessment, one form with two sessions was developed for each grade. The form follows guidelines informed by the test blueprint (the test blueprint is discussed in section 3.4). The operational items are presented in Session 1 and the field-test items are presented in Session 2.

Figure 3-2. Linear Test Design



3.3.2 Operational Items and Field-Test Items

As discussed earlier, there is one form for each grade. As shown in Table 3-4, the test includes 39 operational items in Session 1. The test also includes 9 field-test items in Session 2. The test has 48 items total per grade.

Table 3-4. Operational and Field-Test Items

Grade	Total Operational Items Administered to Each Student	Total Field-Test Items Administered to Each Student	Total Number of Items Administered to Each Student
5	39	9	48
8	39	9	48
HS	39	9	48

The 2023 operational items were selected to adhere to the content blueprints for the test, which has established targets for the percentage of items administered in each science domain (Physical Sciences, Life Sciences, and Earth and Space Sciences).

The 2023 field-test items were selected based on the following criteria:

- they represent a variety of item complexity levels;
- they provide a variety of content; and
- the items are engaging, accurate, and free of bias.

The items on each form are reviewed by Cognia's psychometricians for any validity and reliability concerns. The detailed Test Construction Process document provided the procedures to follow in constructing the test including the psychometric parameters that form the criteria each constructed test should meet. This document is used as the guiding resource to replicate MSAA test construction processes across administration years. The test construction process occurs following data review of the field-test items from the previous administration. Cognia's content specialists create the test forms based on the test blueprints and criteria provided by the psychometricians. The form is then evaluated by the psychometricians and revision loops occur as needed. Once the psychometricians provide approval of a constructed test, it is then also reviewed by the MSAA Science Partners. All constructed tests, as well as the field-test items, are posted on a secure FTP site for the MSAA Science Partners' review. A webinar is held with the MSAA Science Partners to explain the test construction process and to review the Test Construction Design document, which provides information about the items selected. The MSAA Science Partners then have an opportunity to provide input on these decisions.

Following the 2023 test administration, the psychometric team conducted a post-equating check for all science items. Some operational items fell below the psychometric thresholds to be considered acceptable items (see section 9.2 for detail thresholds), presumably due to the newness of the science test and continued pandemic effect. After discussing this with the MSAA Science Partners, a decision was made to use field-test items to replace these operational items for scoring purposes, with the goal to increase the overall reliability of the test. A detailed procedure is documented in Chapter 9.

3.3.3 Item Design

The MSAA Science item design and administration is intended to capture student performance at different levels of skill and knowledge acquisition. The assessment items incorporate important aspects of item design related to both varying levels of content complexity and the degree and type of scaffolds and supports.

The MSAA Science content development process addresses levels of cognitive and language complexity, specifically addressing the content standards, and the heterogeneous characteristics of the target student population. The assessment items vary systematically in complexity yet remain aligned with the EPEs. The items are designed to capture student performance by varying two characteristics: (1) levels of content complexity, and (2) degrees and types of scaffolds and supports. The scaffolds and supports are provided to focus the student on the task and elicit a response without guiding the student's response. The scaffolds and supports are built into the item design for each level of each EPE. The items include the scaffolds and supports that the students are presented with and the DTAs include scripts and specific directions on how the TA is to provide these supports. Building them into the items themselves ensures the supports are being provided consistently to all students who take the test.

The science item specifications for each standard reflect the three item complexity levels. The primary distinctions among items written at item complexity levels 1–3 are (a) their connection to content standards and (b) the scaffolded supports provided at each level.

- Complexity level 3 items target the Level 3 EPE, with minimal supports provided during item administration.
- Complexity level 2 items target the Level 2 EPE, with content supports (e.g., simplified language) provided during item administration.
- Complexity level 1 items target the Level 1 EPE, with content supports (e.g., graphics, simplified language) and item supports (e.g., two response options are non-plausible).

3.3.3.1 Overall Item Structure

A range of item levels is developed for each EPE, as described in section 3.1.1 and the introduction to this section, 3.3.3. Each level provides variable features and supports that offer multiple entry points for a variety of students to demonstrate their knowledge and skill. All items are currently selected-response items worth one point each. Selected-response items consist of a stimulus (optional), a stem, and three options. Details related to each part of the item are outlined in the item specifications.

- Stimulus: The amount of information to be contained in the stimulus/scenario for an item is detailed in the item specifications, specifically in Volume of Information. Level 1 items may or may not contain a stimulus. Level 2 and Level 3 items will almost always contain a stimulus.
- Stem: The item or question stem asks the student a question that elicits a response. Stems are phrased in a way that is open to any response mode so that students can indicate the answer in any way that works for them. Thus, the stem is presented so that it is response-neutral.
- Options: Only one answer option (key) is correct. Options are parallel in construction; for example, equal in level of detail (e.g., graphic options) or equal in length (e.g., sentence/phrases, same number of syllables). Alternatively, if text options are all different lengths, they may be ordered from least to greatest length. Rules for distractor options vary depending on the level of the item.
    - Level 1 items typically will have two irrelevant distractors, meaning that the distractors are completely unrelated to the context presented, or the distractors relate to some element of the stimulus/context but do not include words, pictures, or phrases closely linked to the correct answer. (In rare situations, distractors at Level 1 may be related to the stimulus. These include contexts in which a student is asked to identify a part of a model. In this case, one of the distractors may be a second part of the model that is not the key.

Another exception occurs when a student is asked to distinguish between two opposing conditions, such as sunny and cloudy or dry and wet. One of the distractors can be the condition opposite the key.) Level 1 options will often contain a word plus a graphic to illustrate (often symbolically) the meaning of that word. However, this will not be necessary in all instances.

    - Level 2 items typically will have one irrelevant distractor and one distractor related more closely to the information found in the stimulus. (In rare cases where a data table does not lend itself to an outlier distractor, all response options may come from the data table.)
    - Level 3 items will have two distractors more closely related to information found in the stimulus.

3.3.3.2 Variable Features and Supports

The MSAA Science has a variety of features and supports incorporated into the item specifications for each EPE that allow for multiple entry points and varying degrees of complexity throughout an item set. The three main categories of development that have parameters at each level to ensure a progression of access points include the following:

- Volume of Information (VI): varying amounts of information to be presented in the scenario, ranging from no scenario to complex scenarios that include three or more sentences with extensive information.
- Context: varying degrees of complexity with contexts across item levels, ranging from familiar and immediate settings to unfamiliar and abstract contexts.
- Vocabulary: varying degrees of familiarity with vocabulary presented ranging from very familiar everyday words to complex and abstract content-related vocabulary.

The tables shown in Appendix E provide detailed information specific to each level of these variable features and supports.

3.3.4 Item Components

3.3.4.1 Selected-Response: Science

All directions and materials needed for administering selected-response items are provided in the secure grade-, content-, and form-specific DTA. Selected-response items are presented to students in a standardized and consistent format. Items are presented in two ways:

1. Standalone item sets

    - Each set contains three items (Level 1, Level 2, Level 3) authored to a single EPE progression.
    - Items are independent of one another; each item includes its own stimulus text and optional graphic.
    - Each item is presented in the following order: item stimulus, item stem, three response options.

2. Cluster item sets

    - Each cluster contains one shared stimulus and six items; three items are authored to one EPE progression and three items are authored to a second EPE progression.
    - Items are independent of one another but are all related to the shared stimulus science context.

- Each cluster is presented in the following order: the shared stimulus (text and optional graphics); Level 1, Level 2, Level 3 items authored to the first EPE progression; Level 1, Level 2, Level 3 items authored to the second EPE progression. Each individual item repeats key information and graphics from the shared stimulus, presents the item stem, and then presents three response options.

For all items, students may select a response from the options in a variety of ways (e.g., using the computer mouse, verbalizing, gesturing, using eye gaze or communication devices, using assistive technology). Students' responses are entered into the MSAA test-taking system. If a student has the scribe accommodation, the scribe enters the student-selected response on behalf of the student.

3.4 Content and Blueprints

The test blueprints followed by MSAA Science are consistent with the evidence-centered design process undertaken to develop the summative assessment and with best practices in educational measurement.

Because the *Framework* only defines three or four DCIs in each content domain, and because of the small number of standards being targeted for the MSAA Science (n=12 per grade), four standards per content domain were chosen as the targets for each grade's assessment.

In translating the distribution of content across 13 item sets, however, we could not weight each content domain equally and therefore chose to reflect the content emphases of NGSS and many state *Framework*-based standards for each grade band. The table below shows the content blueprint for the operational test for each grade. The test blueprint for each grade in Appendix F incorporates the overall content distributions used for the development of the operational tests. As noted in Chapter 2, the content distributions in the blueprints were reviewed by various stakeholders and serve as the basis for the current MSAA Science content assessed at each grade level.

Table 3-5. Blueprint for Distribution of Science Content by Grade Level

Science Content Category	Gr 5	Gr 8	HS
Physical Sciences	~40%	30–40%	30–40%
Life Sciences	~30%	30–40%	30–40%
Earth and Space Sciences	~30%	~30%	~30%

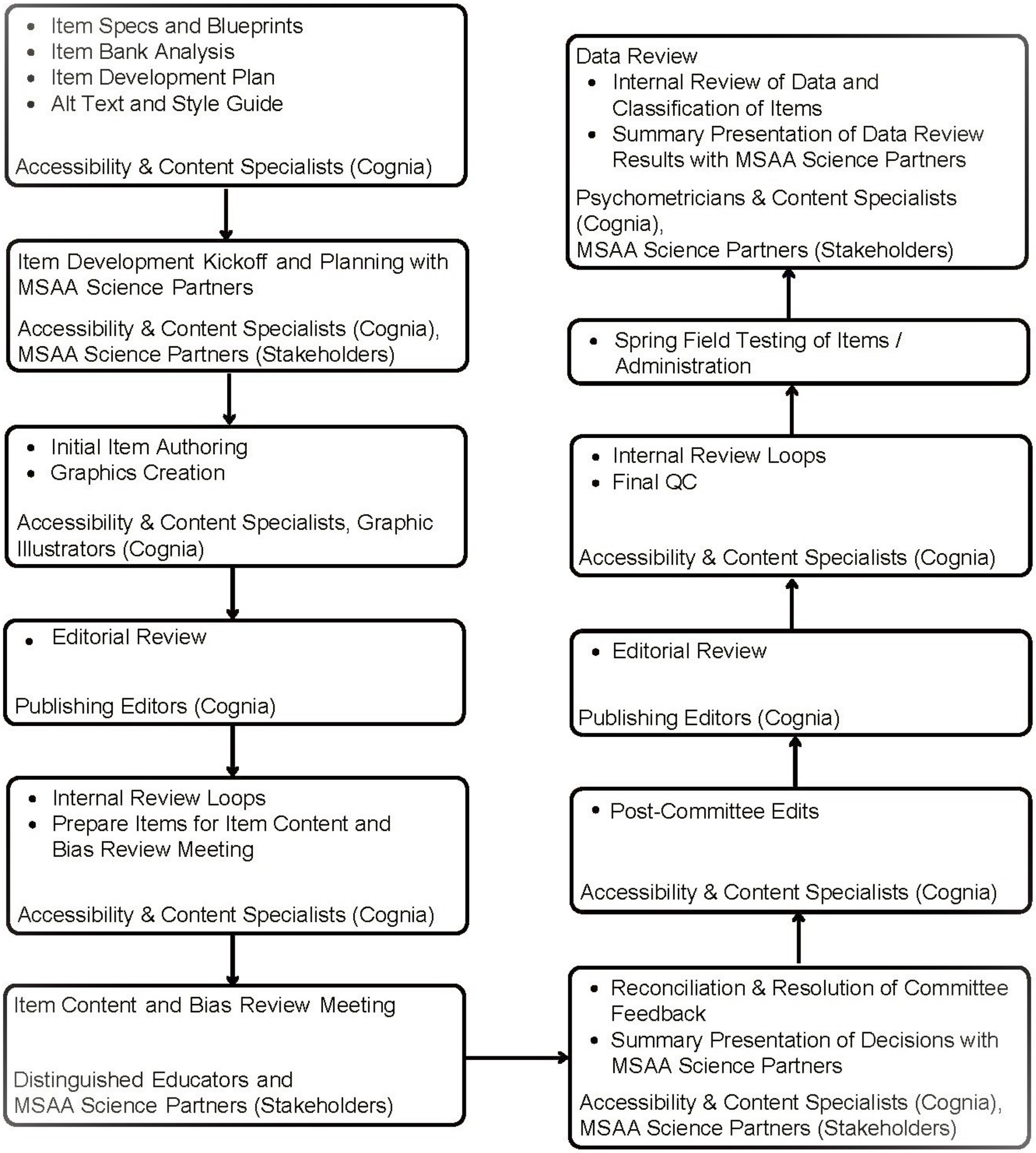
Chapter 4. Test Development: Stakeholder Involvement

4.1 General Philosophy and Role of the MSAA Science Partners and Other Stakeholders in Test Development

As discussed previously, the MSAA Science is part of a comprehensive MSAA program designed to promote increasingly higher academic outcomes for students with the most significant cognitive disabilities in ELA, mathematics, and science in preparation for a broader array of post-secondary outcomes. The MSAA Science is designed to assess the academic content of the EPEs through an assessment design that consists of items written at various levels of complexity and provides built-in supports to meet the individual needs of the students. The assessment allows students to demonstrate what they know and what they can do. Given the wide diversity of the student population, emphasis is placed on ensuring that the MSAA Science is appropriate and accessible to all eligible students.

The MSAA Science items on the 2023 administration are from the previous MSAA Science 2021 field-test administration and the 2022 operational administration. In 2021 the MSAA Science was administered for the first time in Arizona and Maine only, as a field test. The 2023 administration was the second administration of MSAA Science that had operational items. As described in Chapter 3, the items selected as field-test items included in the operational forms are developed by Cognia. The item development process is an iterative one, which allows for multiple opportunities for review of the items by various internal stakeholders including content experts, accessibility specialists, and editors. The items are also reviewed by external stakeholders including the MSAA Science Partners and content experts from the field who participate as panelists in an Item Content and Bias-Sensitivity review meeting. Items that are newly developed are field tested during the spring administration. Once they are field tested, the items undergo data analysis and then go through a data review process with internal stakeholders including psychometricians and content experts as well as the MSAA Science Partners. Figure 4-1 provides a flowchart outlining the item development process.

Figure 4-1. Item Development Process



The Item Content and Bias-Sensitivity review meeting consists of three groups, broken out by grade level, who review the items for content alignment, bias and sensitivity considerations, and accessibility considerations. The list of participants in the Item Content and Bias-Sensitivity review is included in Appendix B. Cognia provides overall direction and guidance regarding field-test item development, and the MSAA Science Partners have an opportunity to provide input during the Item Development Kickoff. This multistage development and review process provides ample opportunity to evaluate items for their accessibility, appropriateness, and adherence to the principles of Universal Design. In this way, accessibility serves as a primary area of consideration throughout the item development process. This focus on accessibility is critical in developing an assessment that allows for the widest range of student participation, as educators seek to provide access to the general education curriculum and foster higher expectations for students with the most significant cognitive disabilities.

Internal stakeholders participate in the data review meeting and are responsible for making determinations about the future usage of the items based on the field-test and operational statistics. During the internal data review meeting(s), Cognia psychometricians, content specialists, and accessibility specialists review the Calibration Report, which includes item statistics for each field-test item that has been flagged by Psychometrics. This year, the operational items were also calibrated and any operational items that were flagged by psychometricians were also reviewed. As shown in Table 4-1, the criteria used for both operational and field-test items are a<=0.25 or SE(b)>0.30. During the internal data review, flagged items are placed into categories. The items might be flagged as Do Not Use (DNU) or flagged as Use with Caution (UWC). The content of the item is reviewed along with the statistics. After each flagged item is reviewed, the internal stakeholders determine whether an item is eligible for usage, rejected (DNU), or designated as revise and re-field test. Rationale is captured for each flagged item as to why the determination was made as such. Additionally, items with strong statistics are reviewed by the content and accessibility specialists with an eye toward what is working for the item and the EPE it is aligned to. This information is used when reviewing the flagged items and making a determination toward designation and is also considered in the new field-test item development process.

Table 4-1. Science Item Evaluation Criteria

Numerical Criteria	Use	Do Not Use (DNU)
Item Difficulty (b)	-4 < b < 4	b < -4 or b > 4
Item Discrimination (a)	a > = 0.25	a < 0.25
Standard Error of Difficulty [SE(b)]	SE(b) < 0.3	SE(b) > =0.3

The items deemed eligible for usage are considered part of the operational item pool and may be selected during the test construction process. The items that are designated as rejected and designated as revise and re-field test do not become part of the operational pool.

The statistics that trigger an item to be flagged are shared with the MSAA Science Partners during a Data Review Summary Meeting. Referenced during data review are the item response theory (IRT) analyses summarized in the Field Test Calibration Report (see section 9.2 for field test calibration details). The MSAA Science Partners are also supplied with Asset Detail Reports, which provide the actual item sets and/or clusters for each of the flagged items. This step allows for the content of the flagged items to be considered when determining future usage.

The MSAA Science Partners also have an opportunity to review and provide input on the constructed tests. As noted previously, this activity occurs following data review. All constructed tests, as well as the field-test items, are posted on a secure FTP site for the MSAA Science Partners to review and provide input. A webinar is held with the MSAA Science Partners to explain the test construction process and to review the MSAA Test Construction Process document, which provides information specific to each content area about the items selected. The MSAA Science Partners then have an opportunity to review and provide input.

Chapter 5. Training and Administration

5.1 Test Administrator and Test Coordinator Training

The MSAA Science Partners adhere to the premise from the testing standards (AERA et al., 2014) that a key consideration in developing test administration procedures and manuals is that test administration should be fair to all examinees. When all Test Administrators (TAs) utilize the same well-defined administration procedures and the provided training, manuals, and supporting documents, administration is prescribed, standardized, and poised to be fair to all examinees. Test Coordinators (TCs) are directly responsible for supporting TAs in understanding and following the administration procedures. Comprehensive TC training and materials targeted to their role and responsibility ensure that they are appropriately prepared to support the TAs.

As explained previously, the MSAA Science is part of the comprehensive MSAA program; as such the training, manuals, and supporting documents provide both comprehensive and content-specific information to TAs and TCs. Given the MSAA Science is a computer-administered test, the administration procedures are consistent with the hardware and software requirements of the test specifications. The MSAA Science requires completion of training by all TCs and TAs to support standardized test processes and procedures. MSAA provides ancillary testing materials each year outlining specific practices and policies including (a) the Test Administration Manual (TAM); (b) MSAA Online Test Administration Training; (c) the MSAA Online Assessment System User Guide for Test Administrators; (d) the MSAA Online Assessment System User Guide for Test Coordinators; and (e) grade-specific Directions for Test Administration (DTA). The online training and the supporting documents are comprehensive and prescriptive, but also provide clear information on where and how much flexibility a TA has while administering the MSAA Science. TCs and TAs receive both the online training and the supporting documents to ensure fidelity of implementation and the validity of the assessment results as well as to help MSAA prevent, detect, and respond to irregularities in academic testing and maintain testing integrity practices for technology-based assessments.

5.2 Test Administrator Training Modules

The online training modules for TAs are available prior to the beginning of the testing window and throughout the testing window. The training modules are customized to address the specific responsibilities of the TA and to provide important information from the three documents TAs are required to use: the (1) TAM, (2) DTA, and (3) MSAA Online Assessment System User Guide for Test Administrators. These training modules were updated for the 2023 administration in correspondence with the updates to the required documents. There are seven modules (see Table 5-1). Each module requires approximately 15–30 minutes to complete.

Table 5-1. Training Modules for Test Administrators

Module 1: MSAA Overview
Module 2: Navigating the MSAA Online Assessment System
Module 3: Test Administrator and Test Coordinator Responsibilities
Module 4: The Writing Prompt
Module 5: Accessibility Features and Accommodations
Module 6: Student Response Check and Early Stopping Rule
Module 7: Science Module

TAs are required to view the training modules (accessed through the MSAA System) in sequence and to successfully complete a final quiz after viewing all modules. Questions pertaining to information in the module follow each online training module for TAs. These questions are included as a review of the content to prepare TAs for the final quiz. TAs must obtain a score of 80% or higher on the final quiz to be certified to access the secure test administration materials. If TAs do not fulfill this certification requirement, they are not allowed access to the secure test materials. The TAs are notified within the MSAA System whether they pass the final quiz. TAs are allowed multiple attempts to obtain a score of 80% or higher on the final quiz.

The Science Module includes additional context about the EPEs, the science item sets and how they align with EPEs, information about clusters, information specific to the layout of the MSAA Science DTA and items, which varies slightly from the ELA and mathematics contents for the MSAA test, and strategies for preparing students prior to testing. The Science Module has quiz questions at the end of the module to assess learning, but there are no questions in the final quiz that pertain to information specific to the Science Module.

In addition to the module training, TAs are instructed to become familiar with the online system by accessing sample items. In addition to the sample items, which were developed by content and measurement experts for teachers, administrators, and policymakers, the MSAA Science added sample items for the 2023 administration that are representative of current MSAA Science item development. The sample items do not address all assessed content at each grade level. Rather, the sample items provide a preview of the array of items and illustrate multiple item features that allow students with a wide range of learner characteristics to interact with the assessment process.

5.3 Test Coordinator Training Modules

Online modules specific to the role of TCs are made available both before and during the testing window. These training modules are customized to address the specific responsibilities of the TCs and to provide important information from the documents TCs are required to use: the (1) TAM and (2) MSAA Online Assessment System User Guide for Test Coordinators. Like the TA training modules, the TC training modules are updated based on the revisions made to the required documents. There are seven modules; each of which runs 20–25 minutes (see Table 5-2).

Table 5-2. Training Modules for Test Coordinators

Module 1: MSAA Overview
Module 2: Navigating the MSAA Online Assessment System
Module 3: Test Administrator and Test Coordinator Responsibilities
Module 4: The Writing Prompt
Module 5: Creating and Managing Users and Classrooms
Module 6: Student Response Check and Early Stopping Rule
Module 7: Science Module

TCs are required to view the online training modules (accessed through the MSAA System) in sequence. Each module must be viewed before the link to the subsequent module becomes accessible. There are quiz questions at the end of each module as a review of the content of that module. TCs are required to complete the online training but not required to take a final quiz.

5.4 Best Practice Videos

The best practice videos are accessed through the MSAA System and provide TAs with targeted information about the MSAA. Video 1 focuses on (1) reviewing assessment features that are available within the MSAA online system, (2) how to go to full screen mode and zoom within the browser, and (3) procedures to follow when using the hybrid approach to administration (i.e., both online and paper-pencil formats). Video 2 focuses on the purpose and steps of conducting the student response check (SRC) and on how to implement the early stopping rule (ESR). (See Table 5-3).

Table 5-3. Best Practice Videos

Video 1: How to Administer an Item
Video 2: How to Administer the SRC and Implement the ESR

5.5 Test Administration Manual

The Test Administration Manual (TAM) provides an overview of, and the guidelines for, planning and managing the MSAA administration for district and school personnel. Additionally, the TAM defines the roles and responsibilities of the TA, TC, and State MSAA Coordinator, all of whom are involved in and oversee the administration of the MSAA. It is organized according to the following tasks:

- providing an overview of the MSAA and the required documents (i.e., TAM, DTA, MSAA Online Assessment System User Guide for Test Administrators, MSAA Online Assessment System User Guide for Test Coordinators);
- defining the roles and responsibilities of the TA and TC, as well as training requirements;
- providing an overview of test designs, item specifics, and content-specific information;
- describing the accessibility features for both online and paper administration as well as the allowable accommodations (i.e., assistive technology, paper version, scribe, sign language); and
- providing detailed information about how to maintain test security and what constitutes a test irregularity.

The TAM also contains a vocabulary list for all content areas, appendices for scribe accommodation and sign language accommodation protocols, the procedures for annotations, and guidelines regarding the use of augmentative and alternative communication by students taking the MSAA. The TAM is accessible to TAs and TCs through the MSAA System and is made available prior to the beginning of the testing window, as well as throughout the testing window.

5.6 Directions for Test Administration (DTA)

The secure grade-, content-, and form-specific DTAs are required to be used by TAs when administering the MSAA Science. Each DTA is accessible through the MSAA System once a TA has been certified. The DTAs are required to be used by the TA for MSAA administration. The elements provided as part of each DTA include standardized directions and scripts that must be followed exactly as written for each item, including alternative text as appropriate.

While the TA has some flexibility in presentation and response mode to ensure the MSAA Science is accessible to a student, the DTAs are designed to provide standardization to ensure a TA is not changing what is being measured.

5.7 Test Coordinator and Test Administrator User Guides

The MSAA Online Assessment System User Guide for Test Coordinators and MSAA Online Assessment System User Guide for Test Administrators provide technical information and troubleshooting tips, plus step-by-step instructions for navigating the MSAA System. Each user guide contains specific information relevant to the role of the TA and the TC. The user guides provide many screenshots that demonstrate the functionality of the MSAA System. The user guides also contain appendices that describe accessibility features, assistive technology compatibility, and the MSAA System technology requirements.

As with the TAM, the user guides are accessible to TAs and TCs through the MSAA System and are available prior to the beginning of the testing window, as well as throughout the testing window.

5.8 Operational Administration

The administration window for the MSAA Science was March 14–April 29, 2023. Regardless of administration format (i.e., online or paper), the student assessments were submitted electronically by the TA on or before April 29, 2023. The MSAA Science is not a timed test. Testing time varies for each student, with testing paused and resumed based on a student’s needs. If a student becomes sick or exhibits frustration, lack of engagement, or refusal to participate during the administration of the MSAA Science, TAs are directed to pause the testing and take a break, which can last for a few minutes or a few days, depending on the student’s needs. The MSAA Science protocols allow the TA to pause and resume the administration of the test as often as necessary during the testing window, based on a student’s needs.

Throughout the administration window, monitoring and quality control processes are ongoing. Support is provided to TCs and TAs through the MSAA Service Center, additional supports built into the MSAA System functionality, and the MSAA Science Partners. TA feedback is gathered through an end of administration test survey. Review of the service center logs, and analysis of the test survey results inform the MSAA Science Partners about areas where clarification and further support is needed.

5.8.1 MSAA Service Center

To provide support to schools before, during, and after testing, Cognia operates and provides tiered technical support through the MSAA Service Center. The MSAA Service Center is available year-round from 6:00 a.m. to 10:00 p.m. EST, Monday through Friday, to accommodate the multiple time zones in which the test is administered.

The TAM directs TAs and TCs to contact the MSAA Service Center with questions pertaining to the MSAA System and test administration procedures. The MSAA Service Center's toll-free support number, e-mail address, and chat link are disseminated to the field through the MSAA System and related communications.

Functionally, support is provided in a tiered manner where Tier 1 support involves direct support to the caller by MSAA Service Center representatives; Tier 2 support includes support by the program management team for items such as policy questions; and Tier 3 support applies to technical requests, which are escalated to the technology vendor for attention.

All activity is tracked in the new MSAA Service Center ticketing system, ServiceNow, and is included in weekly status reports that are provided to MSAA Science Partners. These reports summarize ticket activity, call analysis data (e.g., call duration, hold time), and per-grade/content and per-state/entity test status summaries throughout the administration window.

5.8.2 Additional Supports

In addition to the MSAA Service Center, the Cognia program management team periodically provides direct phone and e-mail support where logistical or procedural support is needed by MSAA state/entity representatives. In cases of Partner specific policy questions, program management refers the individual to the appropriate MSAA Science Partner for support and resolution of the question.

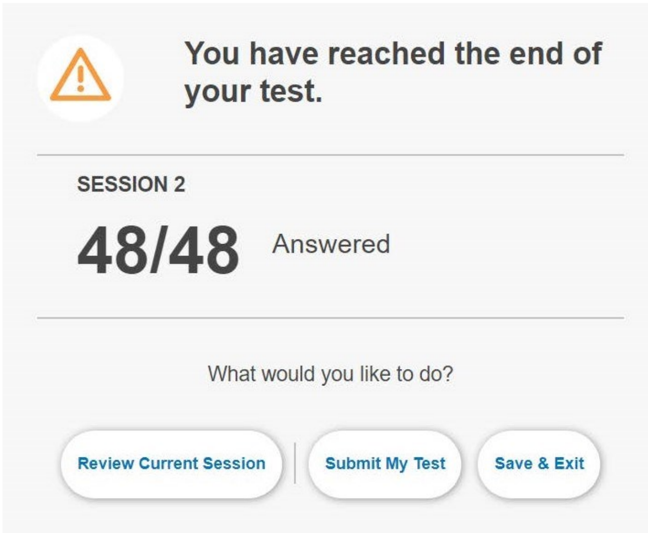
The MSAA System also has a banner messaging system that can be used as needed to communicate updates to the field, whether providing reminders about upcoming important key dates, advanced notice of planned schedule maintenance, or known system issues being experienced. The banner messaging offers color-coding based on severity/importance (e.g., red for high severity; yellow for low severity; and blue for informational). When the messaging system is activated, a banner message appears at the top of the screen upon login to notify users of system information and upcoming system activities, such as known issues and scheduled system maintenance, as well as upcoming test administration deadlines.

5.8.3 Monitoring and Quality Control

To ensure that proper testing procedures and appropriate test practices are maintained throughout administration, numerous measures are taken both to communicate participants' responsibilities and to monitor the appropriateness, accuracy, and completion of key procedures and tasks. The TAM outlines the procedure for reporting any violation or suspected violation of test security or confidentiality by notifying the school or district TC. TCs are then instructed to follow state/entity procedures regarding reporting the issue or suspected issue; however, district TCs are informed that they must report to the State MSAA Coordinator any incidents involving alleged or suspected violations that are considered serious irregularities. The TAM further explains that the consequences for inappropriate test practices are determined by the individual state's/entity's professional codes of ethics and law.

The online MSAA System contains built-in measures to ensure proper testing procedures, as seen in the session-based test design. When the TA clicks the *Next* button on the last question of Session 1, a prompt appears notifying the TA that he or she has reached the end of the session and asks the TA to click Next when ready to start Session 2. When the student responds to the last item in Session 2, a prompt appears notifying the TA that he or she has reached the end of the test, displaying the number of answered items, and allowing the TA to review the current session, submit the test, or save and exit the test.

Figure 5-1. End of Test Prompt



If the TA clicks the *Save & Exit* button, the test will resume the next time on the last item answered. If the TA clicks the *Submit My Test* button, the test is submitted and cannot be re-opened. This prompt reduces the risk of users accidentally submitting a test without properly understanding the implications.

Throughout the administration window, Cognia monitors activity and provides weekly updates to the MSAA Science Partners on the test statuses and on trends identified in support calls. These updates provide a mechanism for concerns to be identified early and the appropriate measures to be taken, such as the creation of assessment-wide or state/entity-level materials and communications. This high level of communication and collaboration throughout the assessment process contributes to a proper and valid administration of the MSAA Science.

5.8.4 Operational Test Survey Results

An End-of-Test Survey (EOTS) allows MSAA to gain knowledge from the experience of each TA administering the test. TAs are instructed to complete at least one EOTS after completing test administration for all their students. The EOTS is administered for all content areas that are assessed. The MSAA Science Partners use the survey to gather feedback from a general program perspective. The survey questions focus on several themes:

- technology use in the classroom,
- student behaviors and engagement,
- instructional time spent on academic content, and
- learning model.

The results of the EOTS highlight several areas of concern that the MSAA Science Partners had identified prior to reviewing the survey data. The data support continued work in the following areas:

- increasing student engagement,
- monitoring the available technology in classrooms to ensure the platform is up-to-date for compatibility, and
- providing professional development to support effective instructional and test-administration strategies.

One issue raised by the teachers in the EOTS data is a lack of continuity between instruction and assessment. The MSAA Science Partners focus on providing professional development to improve instructional practices and to clarify administration policies that increase student engagement by utilizing strategies that align with instruction and still allow for a standard administration.

Several questions on the survey address teachers' viewpoints and philosophies regarding teaching students with the most significant cognitive disabilities. The results again indicate the need for professional development that builds awareness and use of the available instructional and curricular materials, which illustrate various ways that students in this population have access to rigorous academic content.

Furthermore, responses from TAs regarding the high level of difficulty of the test reveal that many students are not fully engaging with the assessment. Individual comments regarding engagement suggest the need for professional development in preparing students for testing. Professional development efforts should make use of the best-practice videos to highlight administration strategies.

The EOTS data also show that many students are using a variety of Augmentative and Alternative Communication (AAC) devices to access the test. In addition, most of the responses indicate that students use desktop computers, laptops, and tablets in the classroom with and without AAC devices and that devices and browsers are compatible with the test. However, some responses indicate that the students in these classrooms either do not utilize or have no access to electronic devices outside of testing. This valuable information can be used to gauge the impact of limited prior exposure to computers on student engagement with the online test.

Chapter 6. Scoring

6.1 Selected-response Item Scoring Processes

6.1.1 Overview of Scoring Process Within the System and Test Administrator Training

Overview of Scoring Process Within the Assessment System

The MSAA System provides automated machine scoring for all the science items. The selected-response item types were described in detail in Chapter 3. The student may provide their responses to the items within the MSAA System. The system also allows for teacher entry of student responses for paper-based test delivery. The selected-response items are scored according to the answer keys provided in each test package.

All item responses are exported from the system and are provided to the Cognia Information Technology Reporting (IT-Reporting) Department. The exported items go through a key verification check to confirm that the selected-response item keys were entered correctly. Items are provided to the content specialists, who conduct a blind key check. Any mismatches to the key recorded in the system are researched by the content specialist, and updates are made following a defined process to update and correct the key. In cases where no mismatches are found, the content specialist notifies the data analyst, and the file is released for final processing.

Items are scored in the MSAA testing system as correct or incorrect, with each of them contributing a score of 1 or 0 to the content-area raw score. Non-responses (blank responses) to any item are scored as 0 points. Detailed score assignments and the comprehensive data analysis requirements are provided in the MSAA Assessments Reporting Services Deliverables Decision Rules document, which can be reviewed in Appendix G.

6.1.2 Test Administrator Training and Support

All TAs must participate in training modules and pass a final quiz to be certified to administer the MSAA, as described in detail in Chapter 5. During the test administration, TAs use the grade-, content-, and form-specific DTAs to administer each item. With the MSAA Science, there are no occurrences where the TA must score an item as they are all selected-response items.

The MSAA Online Assessment System User Guide for Test Administrators provides additional directions to TAs on entering item responses in the MSAA System. The guide outlines the use of the system, including how to enter student responses and submit each content-area test.

For support related to the administration, entry of student responses, and submission of student responses during the administration window, TAs can call or e-mail the MSAA Service Center.

Chapter 7. Reporting

7.1 Development and Approval of Report-Specific Documents

To ensure that reported results for the MSAA Science are accurate relative to collected data, the Reporting Services Deliverables Decision Rules document delineating processing rules is prepared, edited in collaboration with the MSAA Reports Subcommittee, and then approved by all participating MSAA Science Partners prior to processing of the results. The processing and reporting business requirements and participation status structure provide the framework for the reporting requirements, which are defined for each unique report and similarly edited in collaboration with the MSAA Reports Subcommittee. The Reporting Services Deliverables Decision Rules are then approved by the MSAA Reports Subcommittee prior to reporting. This document includes all content areas that are part of the comprehensive MSAA program.

The Reporting Services Deliverables Decision Rules document contains the hierarchy by which the participation statuses are assigned for each individual test, incorporating data elements collected by the test platform and directly from the MSAA Science Partners. The reporting requirements and corresponding report design templates were developed by Cognia with the guidance of the MSAA Reports Subcommittee. Both documents underwent iterative review processes that included draft reviews by the appropriate subcommittee, incorporation of edits, draft reviews by all participating MSAA Science Partners, subcommittee review, and integration of feedback, until final revisions were approved by all participating MSAA Science Partners.

Creating the Report Design Templates

Cognia worked with the MSAA Reports Subcommittee in combination with the MSAA Science Partners to develop the report design templates based on existing report designs that would ensure that the data elements, layout, and report text were meaningful for reporting the spring 2023 MSAA Science results. Once finalized, the results of this collaborative process were presented to participating MSAA Science Partners for final approval.

MSAA 2023 Guide for Score Report Interpretation

Cognia uses an iterative process to annually update the Guide for Score Report Interpretation (see Appendix H) with the MSAA Reports Subcommittee. Updates are made to ensure that the guide provides the most helpful information to district and school staff as they review reports for their own knowledge and as they discuss the reports with parents or guardians. The guide includes an overview of the MSAA, student participation criteria, score reporting overview, and samples of the various types of reports available to schools and districts. Guidelines inform the interpretation and utilization of MSAA scores. The guide also includes explanations for all special reporting codes and messages, as well as performance-level scale score ranges. Partners are permitted to remove codes not used in their state. Appendices included in this guide contain a sample individual student report and the writing prompt scoring rubrics. The final, approved 2023 MSAA Guide for Score Report Interpretation is delivered electronically to the MSAA Science Partners for state-specific revisions and distribution.

7.2 Specific Primary Reports Generated for Schools, Districts, and States

Cognia, in collaboration with the MSAA Reports Subcommittee and the MSAA Science Partners, reviews and updates the following primary reports annually:

- Student reports
- School and district roster reports
- School, district, and state summary reports

Reports are generated for each school, district, or state that has results, as defined by the MSAA processing and reporting business requirements and reporting requirements. These reports, along with student results data files, are posted online via the MSAA Online Assessment System's secure data and reporting portal. As determined by the MSAA Science Partners, only Test Coordinators (TCs) are granted access to the online reports. Access is controlled by user-permissioned accounts, as illustrated in Table 7-1.

Table 7-1. Report/File Availability by Role

		Test Coordinator	
Reports	**State**	**District**	**School**
Student	Yes	Yes	Yes
School Roster	Yes	Yes	Yes
District Roster	Yes	Yes	No
School Summary	Yes	Yes	Yes
District Summary	Yes	Yes	No
State Summary	Yes	No	No
		Test Coordinator	
Data Files	**State**	**District**	**School**
School	Yes	Yes	Yes
District	Yes	Yes	No
State	Yes	No	No

For the purposes of the assessment system, MSAA Science Partners are regarded as State TCs. As such, they can add new district and school TCs to the online system and block from the system any users no longer in the TC role. For 2023, these reports were provided in October to schools, districts, and parents due to the standard setting occurring.

The primary results reported are the student's scale score and performance-level classification for science. The performance-level classifications, with cut scores determined through the original standard-setting process (see Chapter 9 for more information), are reported under the labels Level 1, Level 2, Level 3, and Level 4. Level 4 is the highest attainable performance level.

The average scale score and the percentage of students in each performance level are summarized by school, district, and state on both the roster and summary reports. These summaries allow for the comparison of individual student performance to overall state performance, as well as comparison of school and district results with the overall state results.

7.2.1 Student Report

The student report is a two-sided, single-page document generated for each student eligible to receive a performance level, as defined by the student report requirements. The report contains science results and was developed for parents and guardians of students who participated in the MSAA Science. Reports are organized by school and posted via the secure-access portal for authorized users to download, print, and disseminate to parents and guardians. Each report contains the student name, test grade, and school on the front of the report. The back page contains the student name, state student ID, school, and test grade. Sample student reports are included in the *MSAA 2023 Guide for Score Report Interpretation*.

Page 1 of the report contains the scale score, performance level, and associated performance-level descriptor for the level obtained by the student. A sentence below the graphical display explains the standard error of measurement (SEM) in an easy-to-understand manner by providing the expected range of scores the student would likely earn if tested again.

Page 2 contains a brief overview of the MSAA Science, including examples of some of the built-in supports available during testing, and highlights the compatibility of the assessment with various modes of communication. Parents and guardians are encouraged to discuss with their child's teacher the supports their child used on the MSAA Science.

Tests for students unable to show a consistent observable mode of communication are closed using the Early Stopping Rule, and the lowest scale score is assigned and displayed along with the Level 1 performance level. This is annotated, and in place of the Level 1 performance-level descriptor, the following text is displayed: *Your child did not show a consistent observable mode of communication during the test, and the test was closed by the teacher. Since your child did not complete the test, the results may not be an accurate representation of your child's skills. If you have additional questions, please contact your child's teacher.*

7.2.2 School Roster Report

The school roster report is organized at the school level and provides a by-grade list of all students enrolled in the MSAA Science, with a snapshot of their participation/test status and results. The number of tested students, the average scale score, and the percentage of students at each performance level are summarized for the school, district, and state at the top of the roster. The processing and reporting business requirements and roster report requirements identify which of the participation status codes are included on the roster and which of the participation test status codes are included in each calculation.

The summary information at the top of the school roster report supports interpretation of results by users, typically those at the school and district levels. Given that many schools have a relatively small number of students in this population, the MSAA Science Partners do not suppress information when the number of students participating is small. This practice places an added responsibility on users to understand the data in the context of small numbers and to use all the provided information to understand the results, as explained in the *MSAA 2023 Guide for Score Report Interpretation*.

Student results are listed below the summary section and are identified by name and by state student identification number. It is intended that these data points be used in conjunction with the *MSAA 2023 Guide for Score Report Interpretation*. The following student-level elements are reported:

- Participation/Test Status
- State Compare (comparison to state average)
- Scale Score
- Performance Level

7.2.3 Summary Reports

Summary reports for the MSAA Science are organized at the school, district, and state levels for each entity with at least one student included in summary report calculations. Inclusion in these calculations is defined by the processing and reporting business requirements and summary report requirements. The following information is summarized by grade and displayed for the school, district, and state based on the level of the report:

- Enrolled (number of students enrolled)
- Tested (number of valid student tests)
- Did Not Test (number of enrolled students who did not test)
- Average Scale Score
- Performance Level (number and percentage at each performance level by grade in the state, district, and school)

This summary provides a comparative snapshot of results and participation information at a high level and includes both participation and performance summary information, allowing users to evaluate both aspects of their assessment results as guided by the *MSAA 2023 Guide for Score Report Interpretation*.

7.2.4 Quality Assurance

Quality-assurance measures at Cognia are embedded throughout the entire process of data capture, analysis, and reporting. The data processors and data analysts who work on the project implement quality-control checks of their respective computer programs. Moreover, when data are handed off to different teams within the IT-Reporting Department, the sending team verifies that the data are accurate prior to handoff. Additionally, when a team receives a data set, the first step is to verify the data for accuracy.

A second level of quality-assurance measurement is parallel processing. One data analyst is responsible for writing all programs required to populate the student and aggregate reporting tables for the administration. Each reporting table is assigned to another data analyst on staff who uses the processing and reporting business requirements to independently program the reporting table. The production and quality-assurance tables are compared, and only after 100% agreement is attained are the tables released for report generation.

The third aspect of quality control at Cognia involves the Software Quality Assurance (SQA) team, which works together with the data processing and data analysis teams to ensure quality data are captured and delivered accurately. Quality control checks are being performed by the data processors and data analysts as the data are handed off via multiple internal software tools. These quality checks initialize the accuracy of the data being ingested into the database and subsequent tables/columns. SQA develops a test plan that includes previously agreed upon report designs and decision rule documents. Test cases housed in an internal test cases repository are then executed in a process including, but not limited to, the following steps:

1. Testing data counts of data imported.
2. Testing data quality of individual fields for valid values, such as Gender, Ethnicity, etc.
3. Validating scripts developed by the software developers to ensure that they match business requirements and technical specifications.

In this testing effort to ensure the quality of the data, the SQA team uses a sample of schools and districts selected based on multiple criteria, such as:

- Unique student testing records
- Students complete testing
- Students partially completed testing
- Invalidated students

Working with the data processing and data analysis teams allows for timely and precise turnaround if any data anomalies are found. To allow full transparency and cohesive teamwork in data validation, test cases are tied to tickets outlining required work.

Finally, the SQA team executes test cases validating student printed reports in comparison to the previously agreed-to report design specifications. Once all the test cases have passed, the SQA team notifies the Cognia State Services team for final sign-off and communication.

Additionally, Breakthrough Technologies (BT), our partner vendor, has a designated QA team that assists with ensuring testing and reporting data are accurate. It starts with the BT team performing QA validations on the CBT extracts that are handed off and used for reporting. If needed, there is back and forth between Cognia DP and BT to investigate and resolve any anomalies seen in the data. Once the Cognia Reporting team has completed the reporting cycle and produced all report deliverables, they are handed off to BT via Cognia SFTP site. BT completes a roll up of files creating school, district, and state level zip files. These zip files are posted and available for download in the MSAA System for active Test Coordinator users. BT's QA team does validations on the zip files as well as testing of the platform to ensure user permissions and org hierarchies that are assigned to users are functioning as expected prior to the online reporting window going live in the platform. Partner TCs have a period of time prior to online reporting window opening, where they can access and review their users and take the appropriate action to ensure access is granted to the appropriate people at the appropriate level.

Chapter 8. Preliminary Statistical Analyses

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, particularly racial, ethnic, or gender groups.

Both qualitative and quantitative analyses are conducted to ensure that MSAA Science items meet these standards. Qualitative analyses are described in Chapters 3–6 of this report; this chapter focuses on quantitative evaluations. Statistical evaluations are presented in three parts: (1) classical statistics, (2) differential item functioning (DIF) statistics, and (3) dimensionality analysis of inter-item correlations. The item analyses presented here are based on the administration of the MSAA Science in spring 2023.

8.1 Classical Difficulty and Discrimination Indices

All items are evaluated in terms of item difficulty according to standard classical test theory practices. Difficulty is defined as the average proportion of points achieved on an item and is measured by obtaining the average score on an item and dividing it by the maximum possible score for the item. Selected-response items are scored dichotomously (correct versus incorrect); for these items, the difficulty index is simply the proportion of students who correctly answered the item (*p*-value). A *p*-value index of 0.0 indicates that all students received no credit for the item; an index of 1.0 indicates that all students received full credit for the item.

Items that are answered correctly by almost all students provide little information about differences in student abilities but 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 may indicate knowledge or skills that have not yet been mastered by most students. In general, to provide the best measurement, difficulty indices should range from near-chance performance of 0.33 (for three-option selected-response items) to 0.90, with most items generally falling between approximately 0.3 and 0.7 for all science grades. However, on a standards-referenced assessment, it may be appropriate to include some items with very low or very high item difficulty values to ensure sufficient content coverage. In MSAA Science, IRT item difficulties are employed for evaluating item difficulty, departing from the use of classical item difficulties.

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-test correlation is referred to as the item's discrimination because it indicates the extent to which successful performance on an item discriminates between high and low scores on the test. For selected-response items, the corresponding statistic is commonly referred to as a point-biserial correlation. The theoretical range of these statistics is -1.0 to 1.0, with a typical observed range from 0.2 to 0.7.

A summary of the item difficulty and item discrimination statistics for each grade is presented in Table 8-1. The mean difficulty and discrimination values shown in the table are within typically observed ranges. Please note that grade 8 has only 37 items instead of 39 due to the limitation from the science item pool.

Table 8-1. MSAA Science: Summary of Item Difficulty and Discrimination Statistics by Grade

Grade	Number of Items	*p*-value				Discrimination			
		Min	Max	Mean	SD	Min	Max	Mean	SD
5	39	0.33	0.75	0.52	0.10	0.21	0.55	0.38	0.10
8	37	0.28	0.77	0.50	0.11	0.18	0.55	0.39	0.10
HS	39	0.27	0.67	0.45	0.09	0.16	0.53	0.34	0.09

Note that the classical statistics should be interpreted with caution because the items are three-option selected-response items. Because the items were developed to correspond to different levels, the item statistics have been summarized by level (Table 8-2). Although items tend to be harder for Level 3 than Level 2, and harder for Level 2 than Level 1, the relative difference is much greater when comparing the Level 1 items to the other levels than it is among Levels 2 and 3. Due to the limitation of the science item pool, there were fewer than 13 items, which was the target for grade 8.

Table 8-2. MSAA Science: Item-Level Classical Test Theory Statistics—Summary by Grade and Item Development Level

Grade	Level	Number of Items	*p*-value				Discrimination			
			Min	Max	Mean	SD	Min	Max	Mean	SD
5	1	13	0.49	0.75	0.61	0.06	0.38	0.55	0.46	0.05
	2	13	0.38	0.66	0.50	0.08	0.21	0.49	0.37	0.09
	3	13	0.33	0.55	0.46	0.07	0.21	0.40	0.30	0.06
8	1	13	0.44	0.77	0.59	0.11	0.36	0.55	0.47	0.06
	2	13	0.31	0.64	0.46	0.09	0.23	0.50	0.35	0.09
	3	11	0.28	0.55	0.44	0.07	0.18	0.42	0.32	0.07
HS	1	13	0.40	0.67	0.53	0.08	0.27	0.53	0.41	0.08
	2	13	0.27	0.57	0.43	0.08	0.21	0.44	0.34	0.07
	3	13	0.36	0.47	0.40	0.03	0.16	0.38	0.28	0.07

8.2 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 the 2023 MSAA Science, 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. 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 the 2023 MSAA Science, four subgroup comparisons were evaluated for DIF:

- Male compared with Female
- White compared with Black
- White compared with Hispanic
- 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 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. A list of low DIF items is listed in Appendix I.

The number of items with a “high” DIF index for each level (the cognitive complexity of the item; refer to information in Chapter 3 for further detail regarding the levels) is shown in Table 8-3. Since an item can exhibit DIF for multiple comparisons, an item was counted once if any of the comparisons showed “high” DIF. Table 8-3 shows that only one item was classified as “high” DIF. These results indicate that the item writing procedures were effective. In grade 5, there was one operational item with high DIF. The operational item set was swapped with another field-test item set without high DIF. In high school, there was one operational item with high DIF. This item was swapped with a field-test item set.

Table 8-3. Number of Items with “High” DIF by Level

Grade	Level 1	Level 2	Level 3
5	0	0	1
8	0	0	0
HS	1	0	0

8.3 Dimensionality Analysis

Because tests are constructed with multiple content-area subcategories, and their associated knowledge and skills, the potential exists for several dimensions being invoked beyond the common primary dimension. Generally, the subcategories are highly correlated with each other; therefore, the primary dimension they share typically explains an overwhelming majority of variance in test scores. In fact, 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 calibrating, linking, scaling, and equating the MSAA operational tests.

The purpose of dimensionality analysis is to study test item responses for evidence of violations of test unidimensionality and, if such evidence is found, to understand what it is telling us about possible multidimensionality. In practice, the most common approach is to look for statistically significant violations of local independence (LI), also known as local item dependence (LID). Because LID (i.e., violations of LI) can occur for reasons other than multidimensionality, if evidence of LID is found, the next step is to study the LID to determine its source (or sources), including the possibility of multidimensionality. Hence, we first conducted hypothesis tests to detect statistically significant LID. If it was found, we: (a) estimated the size of the LID and (b) studied the nature of the LID with particular emphasis on possible multidimensionality. Findings from dimensionality analyses performed on the 2021–22 MSAA operational science items 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 et al., 2001) and DETECT (Zhang & Stout, 1999). Both methods use as their basic statistical building block the estimated average conditional covariances for item pairs. 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 of zero, indicating statistically independent item responses for examinees with equal expected total test scores. Nonzero conditional covariances are essentially evidence of LID, which often implies multidimensionality. Thus, nonrandom patterns of positive and negative conditional covariances are indicative of LID, which may imply multidimensionality.

DIMTEST is a hypothesis-testing procedure for detecting LID. 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 LID. The cross-validation sample is then used to test whether the conditional covariances of the selected cluster of items displays LID, 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 for the size of the LID. 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 size of the LID for an item pair. DETECT values less than 0.2 indicate very weak LID (near unidimensionality); values of 0.2 to 0.4, weak to moderate LID; values of 0.4 to 1.0, moderate to strong LID; and values greater than 1.0, very strong LID (Roussos & Ozbek, 2006).

Note that the goal of the dimensionality analysis is to evaluate the assumption of unidimensionality in the IRT model used for the calibration. A procedure was introduced in response to the repeated finding in dimensionality analyses from other MSAA assessments that a small (but nontrivial) percentage of the students, referred to as “R9-stringers,” were exhibiting response behavior incompatible with the assumptions of the psychometric model. R9-stringers are students who respond to nine (or more) consecutive multiple-choice items with the exact same option. Because the calibration data had the R9-stringers removed, the data used in the dimensionality analysis also had the R9-stringers removed. For

2023, the data from R9-stringers were first identified and removed prior to conducting the dimensionality analyses. Table 8-4 summarizes the dimensionality analysis sample sizes both prior to and after removing the data from the R9-stringers.

Table 8-4. Summary of 2022–23 Testing Population

Grade	Total Before Removing Stringers	Total After Removing Stringers	Number of Stringers	Percent Stringers
5	998	799	199	20
8	947	766	181	19
HS	836	672	164	20

DIMTEST and DETECT were separately applied to each grade on the 2021–22 MSAA Science tests. Thus, a total of three analyses were conducted. Within each analysis, each dataset was split into a training sample and a cross-validation sample. The sample sizes across the three analyses varied slightly from a low of 638 in high school to 789 in grade 8.

DIMTEST was then applied to every dataset. Even though the sample sizes were not large for the MSAA Science, the DIMTEST null hypothesis was rejected at a significance level of 0.05 for every dataset. Next, DETECT was used to estimate the LID effect size for all the tests. Table 8-5 displays the LID effect size estimates from DETECT.

Table 8-5. Average Multidimensional Effect Sizes

Grade	DETECT Effect Size without Stringers	
	2021–2022	2022–2023
5	0.81	0.52
8	0.55	0.73
HS	0.41	0.64
Average	**0.59**	**0.63**

The results for 2022–23 displayed in Table 8-5 show that grade 8 and high school showed a moderate level of LID. The DETECT indices for grade 5 indicate a moderate to strong value of LID. Next, an investigation was conducted to identify the possible source(s) of the violations of local independence that could help explain the DIMTEST and DETECT results. For comparison purposes, Table 8-5 also provides the DETECT indices with data that contain stringers. Very strong LID indices were detected in the analyses with stringers, which indicates that the presence of stringers contributes greatly to the magnitude of the LID in the dimensionality analyses.

The removal of the R9-stringers from the data did not eliminate the key-option clustering, but it did greatly weaken its effect. Although efforts were made to identify stringers, there are still stringers out there that were not detected. For example, other stringers patterns, such as R8 or R7 stringers, exist in the student response data that were not detected by the R9-stringers identification business rule. These stringers may contribute to the high values in effect sizes. Additionally, in the test construction, additional efforts were made to carefully limit the number of items in a row having the same key.

8.4 Internal Consistency of Domain Scores

Although the spring 2023 MSAA Science did not report domain scores, examining the internal consistency of the domain raw scores provides additional information on how the test functions as a whole. When domain scores are strongly related to each other, it implies a high internal consistency between domains. The Pearson correlations are shown in Table 8-6. Results generally indicate that the domain scores correlate well with one another and with overall total scores.

Table 8-6. Pearson Correlation of Total Test and Domain Raw Scores on the MSAA Science per Grade

Subtest	Number of Items	Total Test	1	2	3
Grade 5					
Total Test		1	--	--	--
1. Earth and Space Sciences		0.907	1	--	--
2. Life Sciences		0.930	0.769	1	--
3. Physical Sciences		0.939	0.776	0.809	1
Grade 8					
Total Test		1	--	--	--
1. Earth and Space Sciences		0.855	1	--	--
2. Life Sciences		0.935	0.710	1	--
3. Physical Sciences		0.957	0.736	0.846	1
Grade HS					
Total Test		1	--	--	--
1. Earth and Space Sciences		0.893	1	--	--
2. Life Sciences		0.945	0.744	1	--
3. Physical Sciences		0.887	0.718	0.772	1

Chapter 9. Item Response Theory Scaling and Equating

This chapter describes the procedures used to calibrate and scale the 2023 MSAA Science Test. Throughout these psychometric analyses, several quality-control procedures and checks on the processes were implemented. These procedures included evaluation of item parameters and their standard errors for reasonableness, examination of test characteristic curves (TCCs) and test information functions (TIFs) for reasonableness, evaluation of model fit, and evaluation of the scaling results (e.g., parallel processing by the Data and Reporting Services and the Psychometrics and Research Departments).

9.1 Item Response Theory

All MSAA items were calibrated using item response theory (IRT). IRT uses a mathematical model to define a relationship between an unobserved measure of student performance, usually referred to as theta (θ) or student ability, and the probability ($P(\theta)$) of obtaining a particular score on an item. This mathematical relationship is referred to as the item characteristic curve (ICC). In IRT, all items are assumed to be unique 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 $P(\theta)$ (Hambleton & Swaminathan, 1985; Hambleton & van der Linden, 1997). The process of estimating the specific mathematical relationship between θ and $P(\theta)$ is called *item calibration*. After items are calibrated, they are defined by a set of parameters that specify a nonlinear relationship between θ and $P(\theta)$. Once the item parameters are known, an estimate of θ for each student can be calculated based on the student's observed responses to the items. This estimate, $\hat{\theta}$, is considered to be an estimate of the student's true score or an underlying latent variable that best characterizes a student's observed test performance. It has characteristics that may be preferable to those of raw scores for equating purposes because it specifically models examinee responses at the item level and facilitates equating to an IRT-based item pool (Kolen & Brennan, 2014).

For the 2023 MSAA Science tests, the two-parameter logistic (2PL) model was used to estimate the ICC for dichotomous items. The 2PL model for dichotomous items can be defined as:

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

where
U indexes the scored response on an item,
i indexes the items,
j indexes students,
a represents item discrimination,
b represents item difficulty,
θ is the student proficiency, and
D is a normalizing constant equal to 1.701.

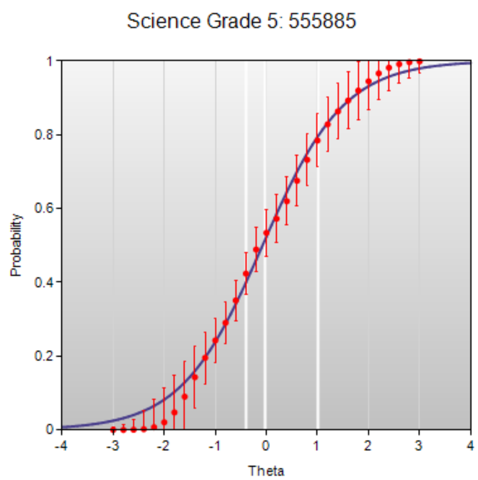
For more information about item calibration and estimation, refer to Lord and Novick (1968), Hambleton and Swaminathan (1985), or Baker and Kim (2004).

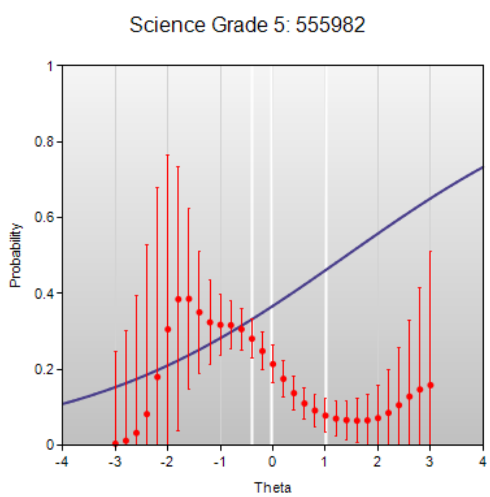
9.2 Calibration Procedure

Because the 2022 MSAA Science was the first operational year of the assessment program, the item parameters for 2022 operational administration were used to establish the new scale. The 2023 science items were calibrated and converted to the 2022 scale. The procedures used to conduct the calibrations are described in this section.

As described in section 8-2, in preparation for the operational and field-test calibrations, the R9-stringers were removed from the data. In calibrating the operational items, first, an un-anchored calibration was conducted on all the operational items using PARSCALE (Muraki & Bock, 2003). At this point, each item was carefully examined for model fit. A visual inspection of the item fit plots was conducted. The empirical proportions of correct responses at a given level of ability must follow the shape of the model-based curve (Figure 9-1).

Figure 9-1. Item Calibration Example Plots (Exemplar and DNU items)





In addition, the item parameter estimates were inspected. Items that violate any of these conditions are automatically marked as “Do Not Use,” though on rare occasions usage is allowed if the item is very close to meeting all the criteria, or is needed for meeting substantive content constraints, and does not result in future tests having poorer reliability than the previous year.

- The discrimination parameters should not be extreme in either direction (neither greater than 3 nor less than 0.25);
- The difficulty parameters should also not be extreme (between -4 and 4);
- Standard Error of difficulty parameter should be less than 0.3.

Next, the field-test items were calibrated using the Fixed Common Item Parameter (FCIP) approach, where the operational items have their item parameters fixed while the field-test items are calibrated. Then the field-test items were evaluated, based on model-fit and item parameter estimates, in the same way as described above for the operational items. Based on the evaluation of model-fit and item parameter estimates, the field-test items were classified as Do Not Use (DNU) if any model-fit issues were identified or if any item parameter estimates fell outside of the criteria. Items that were not classified as DNU were considered eligible and were then uploaded to the item bank.

After examining the operational items, the psychometric team discovered that item statistics of several operational items fell below the psychometric threshold described above. Since 2023 fell within the first three years of the new science program, and given all students take the field-test items, a decision was made in consultation with Cognia content specialists and with the MSAA Science Partners to use field-test items to replace these operational items while maintaining the original content blueprint as much as possible. Table 9-1 lists the number of operational items that were replaced by field-test items. Science items exist in clusters, and if one item within a cluster is deemed Do Not Use (DNU), the entire cluster generally requires replacement. For instance, six items were replaced from the high school science test due to two DNU items identified in separate clusters.

Table 9-1. Summary of Operational Item Changes

Grade	Number of OP Items Removed	Number of FT Items Replacing OP Items	Original Test Length	Updated Test Length
5	3	3	39	39
8	5	3	39	37
HS	6	6	39	39

9.3 Item Response Theory Results

The tables in Appendix J give the IRT item parameters for all the operational items on the 2023 MSAA Science tests by grade based on their pre-equated calibrations. The statistics for the operational items are summarized in Tables 9-2 and 9-3. The mean item parameter estimates shown in the tables are within the generally acceptable and expected ranges. For ease of reference, Table 9-2 displays the means and standard deviations averaged across all dichotomously scored operational items for each grade. To elevate the visual clarity of the data portrayed in Table 9-2, Figure 9-3 presents the same information through boxplots. These visual representations highlight the differences in item difficulties across three item levels.

Table 9-2. IRT Summary Statistics for Dichotomously Scored Items

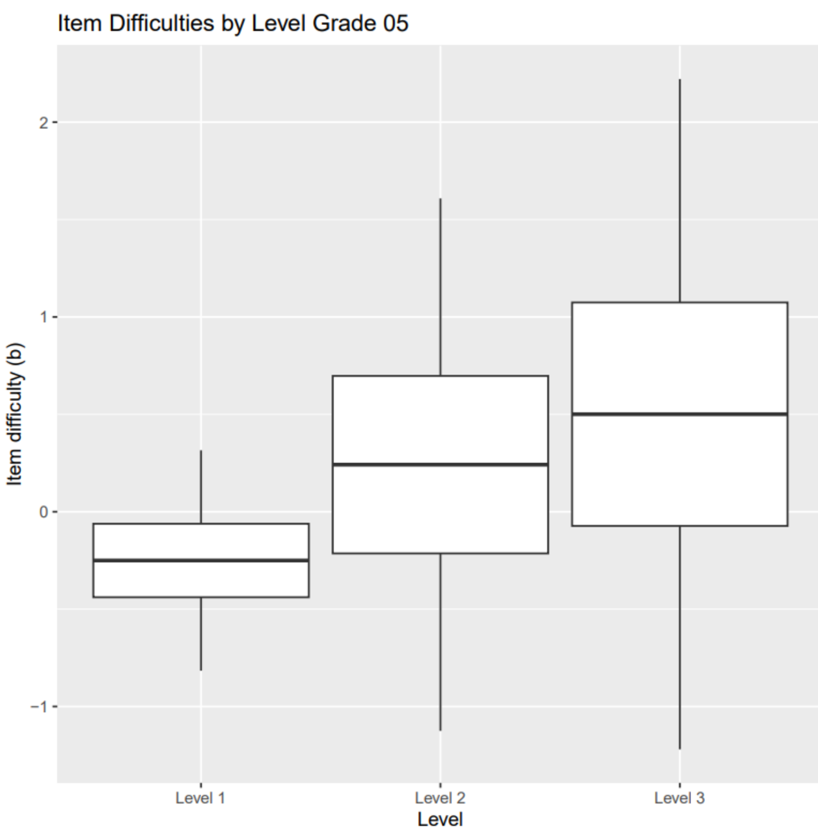
Grade	Number of Items	*a*		*b*	
		Mean	SD	Mean	SD
5	39	0.62	0.29	0.16	0.53
8	37	0.58	0.31	0.30	0.70
HS	39	0.58	0.26	0.45	0.51

Because the items were developed to correspond to different levels, the item statistics have also been summarized by level in Table 9-3.

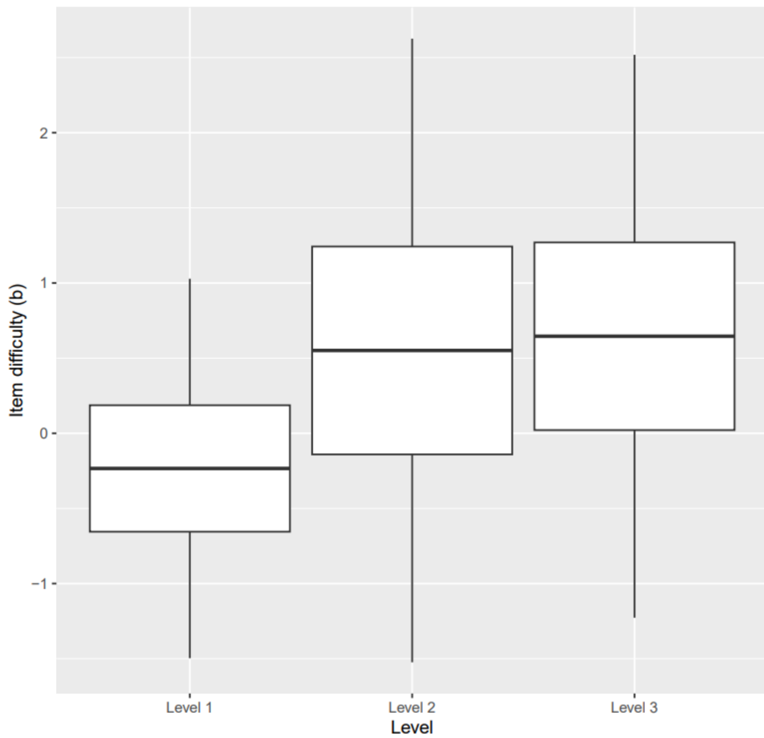
Table 9-3. IRT Summary Statistics by Grade and Level—Dichotomous Items

Grade	Level	Number of Items	*a*		*b*	
			Mean	SD	Mean	SD
5	1	13	0.89	0.28	-0.25	0.19
	2	13	0.57	0.20	0.24	0.46
	3	13	0.41	0.11	0.50	0.57
8	1	13	0.86	0.32	-0.23	0.42
	2	13	0.45	0.18	0.55	0.69
	3	11	0.38	0.10	0.65	0.62
HS	1	13	0.79	0.32	0.04	0.34
	2	13	0.54	0.16	0.50	0.49
	3	13	0.41	0.11	0.82	0.38

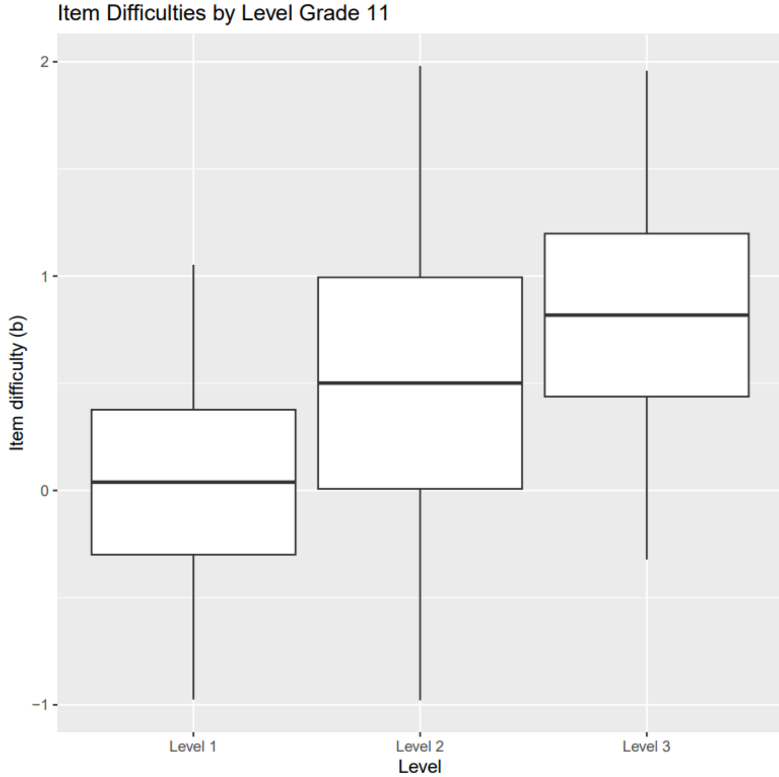
Figure 9-2. Boxplots of IRT Summary Statistics by Grade and Level



Item Difficulties by Level Grade 08



Item Difficulties by Level Grade 11



As seen in Tables 9-1 and 9-2, item difficulty tends to have a positive relationship with level: as the level increases, the items tend to be more difficult (as intended). In nearly all cases, the average difficulty increased from Level 1 to Level 2 and from Level 2 to Level 3. The largest differences were clearly the Level 1 to Level 2 differences for all grade levels. To investigate these tendencies more comprehensively, a one-way analysis of variance (ANOVA) was conducted on item difficulty with level as the factor. Item difficulty differed significantly by grade level.

The ANOVAs indicated that level was statistically significant with an R-squared value of 41%. Tukey paired-comparison tests were also conducted across all grades. These results showed statistically significant differences between Level 1 and each of the other levels. The Tukey comparison for Level 2 versus Level 3 was not significant.

In terms of the difficulty of the whole test, TCCs provide a more complete picture of the relationship between thetas and test scores. The TCCs display the expected (average) raw score associated with each θ_j value between -4.0 and 4.0. Mathematically, the TCC is computed by summing the expected score on all the ICCs of all items that contribute to the raw score. Using the notation introduced in the previous section, the expected raw score at a given value of θ_j is

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

where
X indexes total raw test score,
U_i indexes the scored response on an item,
i indexes the items (and n is the number of items contributing to the raw score),
j indexes students (here, θ_j runs from -2 to 2), and

$E(X|\theta_j)$ is the expected raw score on the test for a student of ability θ_j.

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.

The TIF, $I(\theta)$ (see Lord, 1980, for theoretical definitions and examples of equations), displays the amount of statistical information the test provides at each value of θ_j. Information functions depict test precision across the entire latent trait continuum. There is an inverse relationship between the information of a test and its standard error of measurement (SEM). The SEM at a given θ_j is approximately equal to the inverse of the square root of the statistical information at θ_j (Hambleton, Swaminathan, & Rogers, 1991), as shown below.

$$SEM(\theta_j) = \frac{1}{\sqrt{I(\theta_j)}}$$

Appendix K shows graphs of the TCCs and TIFs for each grade of the MSAA Science. Table 9-4 shows the TCC and TIF values at three cuts per grade. The TIF values, representing the amount of test information at each cut point for different grades, play a crucial role in the test construction process. Ideally, TIF values above 10 are targeted. Table 9-4 reveals that for grades 5 and 11, TIF values surpass 10 for both Level 2 and Level 3 cuts. In the case of grade 8, TIF values exceed 10 at the Level 2 cut and are close to 10 at the Level 3 cut (TIF at cut 3=9.7). However, for Level 4 cuts, TIF values are comparatively lower, ranging from 5 to 9.24 across all grades. This indicates a notable area for ongoing improvement. Given the relatively new nature of science assessments and the ongoing field testing, addressing and enhancing TIF values at the Level 3 cut is essential for future test development efforts.

Table 9-4. TCC and TIF Summary at Three Cut Scores

Grade	Cut Score	TCC	TIF
5	Level 2 Cut	15.63	12.56
	Level 3 Cut	18.96	12.06
	Level 4 Cut	24.12	9.24
8	Level 2 Cut	14.85	10.59
	Level 3 Cut	17.59	9.70
	Level 4 Cut	25.30	5.00
11	Level 2 Cut	15.39	10.54
	Level 3 Cut	17.91	10.18
	Level 4 Cut	24.67	7.23

9.4 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, as well as to equate one year's forms to those given in the previous year. Equating ensures that students are not advantaged or disadvantaged because the test form they took is easier or harder than those taken by other students.

In the inaugural year (2021–22) of MSAA Science, the initial scale was set using one form per grade. In the subsequent 2022–23 administration, item calibration utilized common items from the previous year as anchor items, employing Stocking-Lord transformation to estimate the transformation constants. Consequently, all items in the 2023 administration were updated to the same scale as 2022 using Stocking-Lord transformation constants. In the MSAA Science equating design, all operational items have successfully passed the field test, establishing them as reliable and suitable common items for equating.

As the scale was initially established in 2021–22, and 2023 represents another year for finalizing the scale initiated in 2022, the newly established cut scores were applied to update the initial scale to produce student scale scores in score reports.

9.5 Reported Scale Scores

Because the θ scale used in IRT calibrations is not intuitively meaningful to stakeholders, reporting scales were developed for the MSAA consistent with established professional practice. The reporting scales are simple linear transformations of the underlying θ scale. The reporting scales range from 1200 through 1290 for all grade/content-area combinations. The second cut was fixed at the July 2023 standard setting to be 1240 for each grade level and the theta values associated with 4 are fixed at 1290(HOSS), as evidenced in Table 9-5.

By providing more specific information about the position of a student's results, scale scores supplement performance-level scores. Students' raw scores (i.e., total number of points) on the 2023 MSAA Science tests were translated to scale scores using a data analysis process called scaling, which simply converts from one scale to another. In the same way that a given temperature can be expressed on either Fahrenheit or Celsius scales, or the same distance can be expressed in either miles or kilometers, student scores on the 2023 MSAA Science tests can be expressed in raw scores, or thetas, or scale scores (and linear transformation of the theta metric).

It is important to note that converting from raw scores to scale scores does not change students' performance-level classifications. Scale scores make for more consistent reporting of results. Raw scores are not comparable from year to year because they are affected by the difficulty of the items that appear on each test form. Equating is a statistical procedure that is used to adjust for differences in form difficulty so that scores on alternate forms can be used interchangeably (Kolen & Brennan, 2014). Since the θ scale is used for equating, scale scores are comparable from one year to the next.

The scale scores are obtained by a simple translation of ability estimates ($\hat{\theta}$) using the linear relationship between threshold values on the θ metric and their equivalent values on the scale score metric. Students' ability estimates are based on their raw scores and are found by mapping through the TCC (also known as inversed TCC method). In particular, the bisection method is used to inverse the TCC. The bisection method is a mathematical technique for finding the roots of an equation, in which we refer to the TCC function. It works by repeatedly dividing the theta interval corresponding to a particular value of TCC into smaller intervals. The method is based on the idea that if a continuous TCC changes from a value that is slightly lower than an integer and another value that is slightly higher than that same integer, then the theta value for the integer must be within that interval. For a given TCC function f(x), the Bisection Method algorithm works as follows:

i. Two values a and b are chosen for which *f(a)* is greater than an integer and *f(b)* is smaller than that integer.
ii. interval halving: a midpoint c is calculated as the arithmetic mean between a and b, c = (a + b) / 2. The TCC function *f(x)* is evaluated for the value of c. If *f(c)* equals the desired integer, which means that the theta value for the integer c is found. If *f(c)* is not equal to the desired integer, we proceed as follows:
   - if *f(c)* is greater than the desired integer, then we replace a with c and we keep the same value for b.
   - if *f(c)* is less than the desired integer, we replace b with c and we keep the same value for a. We go back to the previous step and recalculate c with the new value of a or b.

Scale scores are calculated using the following linear equation:

$$SS = m\hat{\theta} + b,$$

where
m is the slope, and
b is the intercept. Specifically, consider the following mathematical expressions:

1. 1240 = m + b* theta (theta at cut 2)
2. 1290 = m + b* 4 (theta at HOSS)

The task is to solve the values of m and b. For MSAA Science, the base-form operational scale was set so that the theta corresponding to the proficient cut from the July 2023 standard setting was transformed to a scale score of 1240. The theta value of 4 was set corresponding to the highest obtainable scale score (HOSS) and transformed to a scale score of 1290. The lowest obtainable scale score (LOSS) was set at 1200. A separate linear transformation is used for each grade and content-area combination. Because only two points within the θ scale score space were fixed (theta at cut 2 and theta at HOSS), the scale score cutpoints between Level 1 and Level 2 and between Level 3 and Level 4 were free to vary across the grade and content-area combinations (as seen in Table 9-4).

Table 9-5 shows the slope and intercept values used to calculate the scale scores for each content area and grade. Note that the values in Table 9-5 will not change unless the standards are reset.

Table 9-5. Scale Score Slope and Intercept by Grade

Grade	Slope	Intercept
5	12.31527094	1240.738916
8	12.27295042	1240.908198
HS	12.84686536	1238.612539

Appendix L contains raw score to scale score lookup tables for the 2023 MSAA Science tests. These are the actual tables used to determine student scale scores, error bands, and provisional performance levels. Graphs of the scale score cumulative frequency distributions for the 2023 MSAA Science tests and for the most recent past test are presented in Appendix M.

9.6 MSAA Science Provisional Performance Levels, Cut Scores, and Standard Setting

Cut scores for MSAA in science were set in a standard-setting process that took place in July 2022 and were finalized in July 2023. Details of the standard-setting procedures can be found in the standard-setting report (Cognia, 2023). A complete description of the standard-setting processes appears in the 2023 *MSAA Science Standard-Setting Report*.

Final cut scores, for the 2023 MSAA Science, appear in Table 9-6a and Table 9-6b.

Table 9-6a. Cut Scores on the Theta Metric and Reporting Scale

Grade	Theta			Scale Score				
	Cut1	Cut2	Cut3	Minimum	Cut1	Cut2	Cut3	Maximum
5	-0.398	-0.06	0.508	1200	1236	1240	1247	1290
8	-0.395	-0.074	1.06	1200	1236	1240	1254	1290
HS	-0.173	0.108	0.926	1200	1236	1240	1251	1290

Table 9-6b. Cut Scores Reporting Scale Range

Grade	Level 1	Level 2	Level 3	Level 4
5	1200–1235	1236–1239	1240–1246	1247–1290
8	1200–1235	1236–1239	1240–1253	1254–1290
HS	1200–1235	1236–1239	1240–1250	1251–1290

Table 9-7 shows the percentage of students by performance-level categories along with the average and standard deviation of the scale scores for each grade/content-area combination. Also, the percentages of

Levels 3 and 4 (levels corresponding to Proficient and above, which are the levels of critical interest for federal accountability purposes) within each grade and content area are provided in the table.

Table 9-7. Percentage of Students by Performance-Level Categories and Scale Score Summary

Grade	Number of Students	Level 1	Level 2	Level 3	Level 4	Levels 3 & 4	Average Scale Score	SD of Scale Score
5	1042	42%	15%	22%	21%	43%	1239	14
8	999	45%	16%	25%	14%	39%	1239	16
HS	911	51%	12%	25%	11%	37%	1235	15

Chapter 10. Reliability and Measurement Error

Tests that function well provide a dependable assessment of the student's level of ability. Unfortunately, no test can do this perfectly. A variety of factors can contribute to a given student's score (as estimated ability) being either higher or lower than their true academic ability. For example, a student may misread an item or mistakenly fill in the wrong bubble when he or she knew the right answer. Collectively, extraneous factors that affect a student's score are referred to as "measurement error." Any assessment includes some amount of measurement error; that is, no measurement is perfect. This is true of all academic assessments—some students will receive scores that underestimate their true ability and other students will receive scores that overestimate their true ability. When tests have a high amount of measurement error, student scores are very unstable. Students with high ability may get low scores or vice versa. Consequently, one cannot reliably estimate a student's true level of ability with such a test. Assessments that have less measurement error (i.e., errors made are small on average and student scores on such a test will consistently represent their ability) are described as "reliable."

There are several ways to estimate test reliability, the most common method is Cronbach's alpha. Cronbach's alpha could be applied to the MSAA Science; however, since the IRT-based reliability was used for MSAA ELA and mathematics, we implemented the same procedure for the MSAA Science for consistency purposes. Thus, an IRT-based estimate of reliability that results in a single value for each grade-level assessment was used.

10.1 IRT Marginal Reliability

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 basic 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 scale CSEM,
$Var\ (\hat{\theta})$ is the variance of theta estimates, and
$Var(SS)$ is the scale score variance.

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 10-1 presents descriptive scale score statistics, IRT-based reliability, and mean scale score CSEMs by grade. (Statistics are based on operational items, which counted toward students' reported scores only.) As shown in the table, most of the values reached levels associated with adequate reliability (0.80 or higher). The mean Scaled CSEM shows that the precision of the scale score is within the normal range.

Table 10-1. IRT Marginal Reliability by Grade

Grade	Min	Max	Mean	SD	IRT Marginal Reliability	Mean Scaled CSEM
5	1201	1290	1242.93	12.48	0.87	4.29
8	1200	1290	1243.17	14.37	0.88	4.81
HS	1200	1290	1240.05	12.14	0.85	4.53

10.2 Subgroup Reliability

The reliability coefficients discussed in the previous section were based on all students who took a MSAA Science test. As an alternate assessment program, it is likely that there are reliability differences across subgroups. For this reason, reliability coefficients for different subgroups were calculated, including gender, ethnicity, limited English proficiency (LEP) status, socioeconomic status, migrant status, and various disability categories. Appendix N presents reliabilities for various subgroups of interest. Subgroup reliabilities were calculated using the IRT-based formula (defined above) based only on the members of the subgroup in question in the computations; values were calculated only for subgroups with 100 or more students and where more than 25% of the students scored above the LOSS (lowest obtainable scale score, which was 1200).

For several reasons, the results relating to subgroup reliability should be interpreted with caution. First, reliabilities are dependent not only on the measurement properties of a test but on the statistical distribution of the studied subgroup. For example, it can readily be seen in Appendix N that subgroup sample sizes varied considerably, which results in a natural variation in reliability coefficients. Alternatively, reliability, which is a type of correlation coefficient, may be artificially depressed for subgroups with little variability (Draper & Smith, 1998). Second, there is no industry standard to interpret the strength of a reliability coefficient, especially when the population of interest is a single subgroup. Again, the reliability statistics provided in the tables in Appendix N should be cautiously interpreted because of the restriction of range mentioned earlier (Section 8.1). The subgroup IRT reliability ranges from 0.79–0.89 for all three grades.

10.3 Reliability of Performance-Level Categorization: Accuracy and Consistency

While related to reliability, the accuracy and consistency of student classification into performance categories are even more important statistics in a standards-based reporting framework (Livingston & Lewis, 1995). After the provisional performance levels were specified and students' performances were classified into those levels, empirical analyses were conducted to determine the statistical accuracy and consistency of the classifications. For the MSAA, students are classified into one of four performance levels: Level 1, Level 2, Level 3, or Level 4. This section of the report explains the methods used to assess the reliability of classification decisions, and results are provided.

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. 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 MSAA 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 whose true and observed classifications matched) signified overall accuracy.

For the classification consistency tables, cell [*i*, *j*] of this table represents the estimated proportion of students whose observed scale score on the first of the two hypothetical parallel tests would fall into classification *i* (where *i* = 1 to 4) and whose observed scale score on the second hypothetical parallel test would fall into classification *j* (where *j* = 1 to 4). The sum of the diagonal entries (i.e., the proportion of students categorized by the two forms into exactly the same classification) signified overall consistency.

Another way to measure consistency is to use Cohen's (1960) coefficient κ (kappa), which assesses 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:

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

where

$C_{i.}$ is the proportion of students whose observed performance level would be Level i (where i = 1–4) on the first hypothetical parallel form of the test;

$C_{.i}$ is the proportion of students whose observed performance level would be Level i (where i = 1–4) on the second hypothetical parallel form of the test; and

C_{ii} is the proportion of students whose observed performance level would be Level i (where i = 1–4) on both hypothetical parallel forms of the test.

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

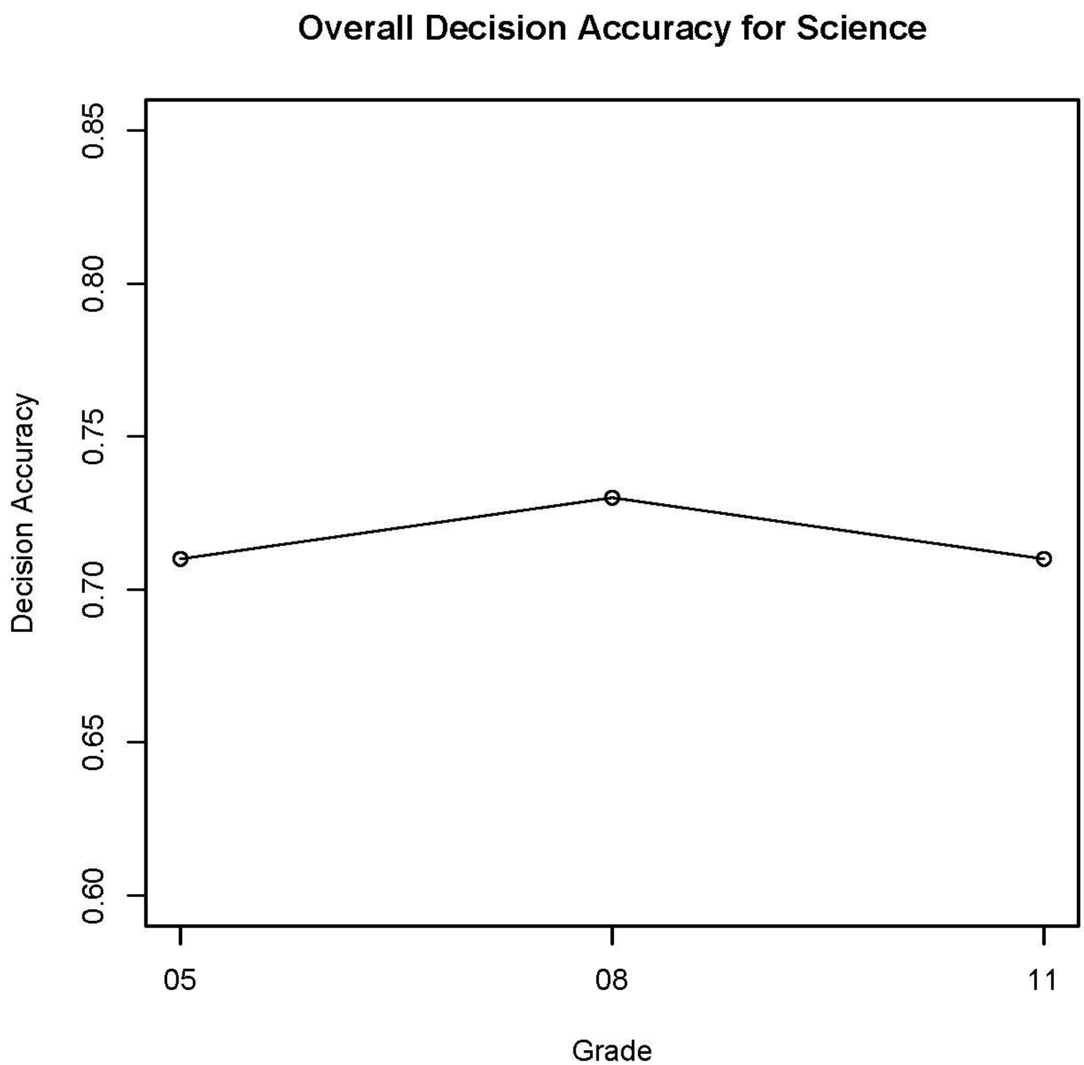
Figure 10-1 shows the overall decision accuracy for science by grade level. More details on decision accuracy and consistency (DAC) are provided in Appendix O. Table O-1 in Appendix O includes overall accuracy and consistency indices, along with kappa. Accuracy and consistency values conditional on performance level are also provided in Table O-1. For these calculations, the denominator is the proportion of students associated with a given performance level. Following is an example from Table O-1, looking at Level 1 for grade 5.

- The conditional *accuracy* value was 0.71. This indicates that among the students whose *true scale scores* placed them in Level 1, 71% would be expected to be in this same level again when categorized according to their observed scale scores.
- The *consistency* value was 0.63. This indicates that among the students whose *observed scale scores* placed them in Level 1, 63% 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 Every Student Succeeds Act (ESSA—see Appendix P for a list of acronyms) accountability purposes, the primary concern is distinguishing between students who are proficient and those who are not yet proficient. For the 2023 MSAA Science, Table O-2 in Appendix O 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.

The overall DAC accuracy for science is lower compared to ELA or mathematics (approximately 0.75 for mathematics and approximately 0.82 for ELA). This difference is partly attributed to ELA/Math being adapted tests with an established item bank.

Figure 10-1. Overall Decision Accuracy for Science by Grade



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

Chapter 11 provides an overview of the primary intended score interpretations and uses of the MSAA Science assessment, including an in-depth review of the assumptions and evidence supporting them. The Standards for Educational and Psychological Testing emphasize the importance of evidence in supporting interpretations and uses of test scores. The goal of any evidence-rating scale presented in this report is to **evaluate the degree to which the evidence supports each score interpretation and use (SIU) claim.** The chapter uses a three-dimensional rating scale to evaluate the evidence supporting each SIU claim, aiming for a comprehensive evaluation (see Appendix Q for additional details concerning the element level rating scale). MSAA Science assessment is part of the comprehensive alternate assessment program, which also includes the MSAA ELA and mathematics. Some ratings of evidence are unique to the MSAA Science, while others are relevant to all three content areas.

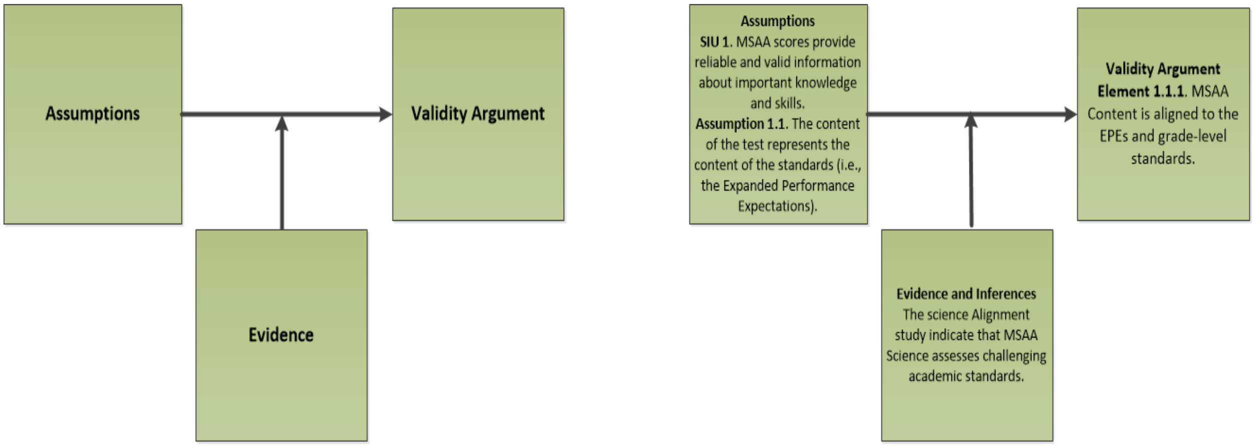
The MSAA Science Validity Argument Logic Model

The MSAA validity argument model involves documenting evidence that connects the assumption-validity argument pairs, as depicted in Figure 11-1. The left-hand panel displays the validity logic model related to MSAA, and the right-hand panel shows an example of how the validity argument is connected to assumptions. Evidence supporting the assumptions is also connected to the assumption-validity argument pairs.

The multi-dimensional rating scale used in the MSAA validity argument model includes three separate scales for assessing the evidence with respect to its relevance, completeness and overall support for a given assumption. The validity model argues that the existing design, procedural evidence, and psychometric evidence support the four intended score interpretations and uses. Each interpretation and use are underpinned by a set of assumptions, which are, in turn, underpinned by elements that require evidence for validation. Detailed information on the four intended score interpretations and uses, the assumptions and elements connecting the evidence to the interpretations and uses, and the supporting evidence is provided in Chapters 2–10, with Table 11-1 summarizing the relationships among the score interpretations and uses, assumptions, and elements.

The rating scale indicates different levels of relevance, completeness, and overall support. Relevance refers to the degree of applicability of the evidence and its ability to withstand challenges, completeness assesses whether all necessary evidence is provided, and overall support evaluates the degree to which the evidence, as a whole, supports the claim. The primary score interpretation and use statements (SIUs) are identified as the main focus of the evidence evaluation.

Figure 11-1. MSAA Validity Argument Model



Adapted from Ferrara & Qunbar (2022) and Chapelle (2021) Figures 2.1-2.3, Kane (2013) Figure 1, and Toulmin (1958).

Table 11-1. Relationships Among Score Interpretations and Uses, Necessary Assumptions, and Elements That Support the Assumptions

Necessary Assumptions		Elements That Support Assumptions
Primary Intended Score Interpretation The MSAA Science scores provide reliable and valid information about important knowledge and skills in elementary, middle, and high school multidimensional science concepts that students with the most significant cognitive disabilities are attaining.		
1.1	The content of the test represents the content of the standards (i.e., the Extended Performance Expectations).	
	1.1.1.	The Extended Performance Expectations are aligned to the standards (aka, Performance Expectations) from *A Framework for K-12 Science Education* and, as such, aligned to each partner's academic content standards for each grade level.
	1.1.2.	The 2023 MSAA Science items are aligned to the Extended Performance Expectations.
	1.1.3.	Science Extended Performance Expectations, which are assessed on the MSAA Science, are aligned with each partner's content standards for each grade level.
	1.1.4.	MSAA Science items are aligned to the MSAA performance level descriptors.
1.2	MSAA Science test items are construct relevant. The elements related to this assumption involve the skills and cognitive processes needed to respond to a specific item, and their alignment with those in the PLDs	
	1.2.1.	Items require application of the KSAs of the targeted construct.
	1.2.2.	Items are accessible to all students, allowing students the opportunity demonstrate what they know and be able to do.
	1.2.3.	Appropriate accommodations are provided to meet student needs.
	1.2.4.	Scaffolding (information provided to vary item difficulty) does not introduce irrelevant variation to the construct.
	1.2.5.	Item rendering (i.e., how items are presented in the testing platform) does not interfere with student access to test content.
	1.2.6.	Test Administration Platform does not interfere with student interaction with test content.

continued

Necessary Assumptions		Elements That Support Assumptions
	1.2.7.	Items are free of bias and sensitivity issues.
1.3	Test administrations in MSAA states followed prescribed, standardized procedural requirements.	
	1.3.1.	Test Administrators and School and District Coordinators understood and performed their roles properly.
	1.3.2.	Test security protocols were diligently followed, and test security concerns and breaches were limited.
1.4	Test scores on the MSAA Science provide reliable information about student performance and accurate classifications into performance levels.	
	1.4.1.	MSAA scores and categorizations into performance levels are adequately reliable for their intended purpose.
	1.4.2.	Item characteristics (i.e., item difficulties) support intended interpretations about all students who take the MSAA.
	1.4.3.	Test characteristics support intended interpretations about all students who take the MSAA.
	1.4.4.	Scaling of the MSAA supports intended interpretations about all students who take the MSAA.
	1.4.5.	Equating of MSAA test forms supports intended interpretations about MSAA students.
1.5	Item and test scoring were implemented accurately.	
	1.5.1.	Machine scored items were scored accurately.
1.6	MSAA Science scores correlate with external indicators of student proficiency (i.e., concurrent and predictive evidence).	
	1.6.1.	MSAA scores correlate as expected with other measures of student proficiency.
Primary Intended Score Use 1 Schools and districts use the MSAA Science and its results to (a) monitor trends in school performance, and (b) design professional development for teachers on how to monitor trends.		
2.1	Schools and districts use the MSAA Science and its results to monitor trends in school performance	
	2.1.1.	MSAA Science scale scores for groups of students are adequately reliable and valid to enable school, district, and state leaders to monitor changes in means, standard deviations, and proficiency level percentages for classroom, school, district, and state groups.
	2.1.2.	MSAA Science 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.
2.2	MSAA Science results are used to design professional development for teachers.	
Primary Intended Score Use 2 The MSAA Science and its results are used to help teachers integrate MSAA Science scores and other information with their instructional planning.		
3.1	Teachers use the MSAA Science and its results to better integrate assessment with their instructional planning.	
	3.1.1.	Teachers find the performance level descriptors and their students' performance levels useful for planning instruction, especially students in performance levels 1 and 2.
	3.1.2.	Teachers find their students' scale score information useful for planning instruction, especially students in levels 1 and 2.
3.2	Teachers use MSAA Science scores and other information for instructional planning.	
Primary Intended Score Use 3 Parents understand and interpret MSAA Science scores and other information correctly to understand what their child knows and can do.		
4.1	Parents find MSAA Science scores and other information useful for understanding what their child knows and can do.	
	4.1.1.	Parents understand and interpret correctly MSAA Science scores and other information to understand what their child knows and can do.

continued

Necessary Assumptions		Elements That Support Assumptions
	4.1.2.	Parents use MSAA Science scores and other information appropriately to understand what their child knows and what their child can do and make decisions about their child's education and learning needs.
4.2	Parents find MSAA Science scores and other information useful for understanding their child's progress over time	
	4.2.1.	Parents understand and interpret correctly MSAA Science scores and other information to understand their child's progress over time.
	4.2.2.	Parents use MSAA Science scores and other information appropriately to understand their child's progress over time and make decisions about their child's education and learning needs.

Relevance of the Evidence

We assess the relevance of each set of evidence provided for every assumption and element, closely aligning with Toulmin and Chapelle's argumentation model, which determines the reliability of the evidence. It's important to note that individual pieces of evidence within a set may vary in their relevance. For example, while test content directly relates to the MSAA Extended Performance Expectations, we lack direct evidence regarding the connection between these EPEs and long-term post-secondary outcomes. MSAA relevance rating scale is as follows:

- **Highly Relevant:** The evidence is strongly connected to the assumption and element.
- **Moderately Relevant:** The evidence provides a noteworthy, though not necessarily strong, connection to the assumption and element.

Completeness of the Evidence

Completeness is defined as having all necessary or appropriate components. The Evidence rating is defined as follows:

- **Complete Evidence:** Includes all relevant evidence in a collection to support a validity argument.
- **Moderate to Substantial Evidence:** Offers several or nearly all relevant pieces of evidence for an assumption/element, though not all required pieces may be available.
- **Limited Evidence:** Comprises only one or two pieces of evidence, which might be marginally relevant, or when more than one or two pieces are needed.
- **No Evidence:** Indicates the absence of any relevant evidence.

Overall Support

Finally, we provide an overall evaluation of the degree to which the collection of evidence supports a claim/assumption/element. This is intended to be a holistic evaluation of the available evidence, rather than a composite of the evaluations in the other two rating scales.

- Evidence **strongly supports** the assumption or element.
- Evidence **moderately supports** the assumption or element.
- Evidence provides **limited support** of the assumption or element.
- Evidence **does not support** the assumption or element, or the evidence does not exist.

The primary score interpretation and use statements (SIUs) for which supporting evidence is needed are as follows.

Primary Intended MSAA Science Score Interpretation

MSAA Science scores provide reliable and valid information about important knowledge and skills in elementary, middle, and high school multidimensional science concepts that students with the most significant cognitive disabilities are attaining.

Primary Intended MSAA Science Score Uses

- Schools and districts use the MSAA Science and its results to (a) monitor trends in student performance and (b) design professional development for teachers.
- Teachers use the MSAA Science and its results to better integrate assessment with their instructional planning.
- Parents use the MSAA Science and its results to get information about (a) what their child knows and what their child can do and (b) their child's progress from year to year.

11.1 Primary Intended Score Interpretation

The MSAA Science scores provide reliable and valid information about important knowledge and skills in elementary, middle, and high school multidimensional science concepts that students with the most significant cognitive disabilities are attaining.

Assumption 1.1. The content of the test represents the content of the standards (i.e., the Extended Performance Expectations).

1.1.1. The Extended Performance Expectations are aligned to the standards (i.e., Performance Expectations) from *A Framework for K-12 Science Education* and, as such, aligned to each partner's academic content standards for each grade level.

1.1.2. The 2023 MSAA Science items are aligned to the Extended Performance Expectations.

1.1.3. Science Extended Performance Expectations, which are assessed on the MSAA Science, are aligned with each partner's academic content standards for each grade level.

1.1.4. MSAA Science items are aligned to the MSAA performance level descriptors.

In 2022, an alignment study was conducted to support the test alignment assumption and its elements for the MSAA Science. The evidence for each element is as follows:

- The Extended Performance Expectations (EPEs) align with the relevant standards and have undergone a thorough development process. Stakeholders reviewed and confirmed domain coverage, content centrality, and performance centrality, with results listed in Chapter 2.
- The 2023 MSAA Science items are aligned with the Extended Performance Expectations. The alignment study conducted in September 2022 confirmed strong alignment for content centrality, performance centrality, domain concurrence, and blueprint representation, with at least 90% of items aligned. In summary, the grade-level items were well-aligned for both content centrality and performance centrality. At least 90% of items were judged as having *some* or *all* the same performance expectations of the EPEs. Domain concurrence for each grade level was well-aligned, at least 90% of the items on the test form align to an EPE defined in the blueprint, no item on the test form reflects expectations not defined in the grade level, and each of the domains in the blueprint is represented by items on the form.
- MSAA Science Partners have adopted academic content standards assessed by the MSAA Science, ensuring alignment with the Science Extended Performance Expectations.

- The 2023 operational MSAA Science items align with the MSAA performance level descriptors through item specifications that provide guidelines for item development. The science items are developed following item specifications for each EPE. The item specifications accomplish two purposes: They (1) provide both general and specific guidelines for developing all test items at the grade levels assessed, and (2) describe the test items and provide samples as reference. Sections in the specification documents are dedicated to information about target EPEs, item contexts, variable features, cognitive task levels, use of graphics, item style and format, and general content limits by academic grade-level content target. As such, because items are developed from the item specifications, they are aligned to the performance level descriptors. There are 12 EPEs, with three levels of items, on the blueprint. Each set of PLDs was written specifically for each EPE (12 sets of PLDs in total). With the limitation of the current item bank, MSAA Science does not have items aligned to each EPE and PLD combination (144 combinations). However, this limitation will be alleviated and addressed in future test development cycles.

Relevance: Evidence is highly relevant.

Completeness: Evidence is complete.

Overall Support: Existing evidence strongly supports the assumption.

Assumption 1.2. MSAA Science test items are construct relevant. The elements corresponding to this assumption are concerned with the skills and cognitive processes required to understand and respond to an item in particular and whether they correspond to the skills and processes required in the PLDs.

1.2.1. Items require application of the KSAs of the targeted construct.
1.2.2. Items are accessible to all students, allowing students the opportunity to demonstrate what they know and are able to do.
1.2.3. Appropriate accommodations are provided to meet student needs.
1.2.4. Scaffolding (information provided to vary item difficulty) does not introduce irrelevant variation to the construct.
1.2.5. Item rendering (i.e., how items are presented in the testing platform) does not interfere with student access to test content.
1.2.6. Test Administration Platform does not interfere with student interaction with test content.
1.2.7. Items are free of bias and sensitivity issues.

The evidence under this assumption is interrelated and is supported by the 2023 alignment study. The assessment content mirrors the content of the standards, which are the Extended Performance Expectations. Additionally, in the science standard setting, a process was included in which subject matter experts evaluated the knowledge, skills, and abilities (KSA) demands of the items relative to the KSAs in the PLDs, which provides additional evidence.

During the item development process, the items followed a rigorous development cycle, including reviews by MSAA Science Partners and by Item Content and Bias and Sensitivity panelists. See Chapter 3 for a detailed description of the item review process.

In differential item functioning (DIF) analyses, we examine subgroup differences in performance when sample sizes permit. Actions are taken to ensure that differences in performance are due to construct-relevant, rather than irrelevant, factors. A detailed description of the DIF analysis procedures is given in

Chapter 8 along with a summary of the results, and additional results presented in Appendix I. Data review also examines the bias and sensitivity of the tested items. Bias and sensitivity checks are also implemented during the item development process. A cognitive lab may also help inform the targeted construct. However, because MSAA Science test development uses the experiences from the MSAA ELA and mathematics interaction study, a separate Cognitive Lab is not planned at this time.

Relevance: Evidence is highly relevant.

Completeness: Evidence is limited.

Overall Support: Existing evidence moderately supports the assumption.

Assumption 1.3. Test administrations in MSAA states in 2023 followed prescribed, standardized procedural requirements.

1.3.1. Test Administrators and School and District Coordinators understood and performed their roles appropriately.
1.3.2. Test security protocols were diligently followed, and test security concerns and breaches were limited.

The evidence shows comprehensive training was provided for Test Administrators and Test Coordinators, ensuring their proper understanding and execution of their roles. This training involved six online modules addressing their specific responsibilities and the required documents. Additionally, a final quiz with an 80% passing score was mandated after completing the modules. Test Coordinators received additional training with access to supporting documents and resources, including best practice videos, a technical support chart, and a survey. Detailed information regarding the procedure and documentation can be found in Chapter 5. Observations in the field were conducted to assess the effective implementation of training, and results indicating 92% compliance with Directions for Test Administration (DTA) instructions and secure storage of materials in 13 observations during the 2023 MSAA Science administration.

Relevance: Evidence is highly relevant.

Completeness: Evidence is complete.

Overall Support: Existing evidence strongly supports the assumption.

Assumption 1.4. Test scores on the 2023 MSAA Science provide reliable information about student performance and accurate classifications into provisional performance levels.

1.4.1. MSAA scores and categorizations into provisional performance levels are adequately reliable for their intended purpose.
1.4.2. Item characteristics support intended interpretations about all students who take the MSAA.
1.4.3. Test characteristics support intended interpretations about all students who take the MSAA.

1.4.4. Scaling of the MSAA supports intended interpretations about all students who take the MSAA.
1.4.5. Equating of MSAA test forms supports intended interpretations about MSAA students.

These elements pertain to the assessment and scaling of the MSAA (Multi-State Alternate Assessment). MSAA assessment process involves extensive evidence collection and analysis to ensure reliability, consistency, and alignment with intended interpretations. While some challenges are identified, they are balanced by the overall strength of the evidence supporting the validity and accuracy of the assessment.

For the internal consistency and reliability of the MSAA tests. The evidence provided demonstrates adequate reliability for all MSAA tests, with an IRT marginal reliability value of 0.8 or higher. The evidence also discussed the standard errors of scaled scores, allowing for the interpretation of test results with consideration of measurement error. Additionally, it presents performance-level classification consistency and accuracy estimates, ensuring the reliability and consistency of classification outcomes.

For the characteristics of individual test items, the evidence described how the item parameters, including discrimination and difficulty levels, are estimated during field testing, using classical, Differential Item Functioning, and IRT analyses. Items that meet statistical and Data Review criteria are selected for operational use. This process ensures that the individual test items align with the intended interpretations for all students.

For the test characteristics, the evidence discussed dimensionality analysis and test information functions. The dimensionality analysis identifies some local item dependence in specific grades, which is attributed to examinee behavior. The Test Information Functions (TIF) are assessed to ensure they align with desired benchmarks around each performance level. A potential challenge related to Cut 3 TIF has been identified, which may be addressed by increasing the item pool in subsequent years.

For the scaling of the MSAA to support intended interpretations about students, various analyses were discussed, including DIF analyses, dimensionality analysis, calibration, and model fit (See Appendix Q for more details). These analyses aim to ensure that the scaling aligns with the constructs intended to be measured and maintains unidimensionality. The evidence suggests that some items exhibit violations of local independence, primarily related to student behavior, but the evidence generally supports strong scaling practices. It also notes that total scores for scaling provide only negligible underestimation of ability standard errors.

Regarding the equating of MSAA test forms to maintain consistency and support the intended interpretations of MSAA students, it's important to note that, as this year represents another standard-setting cycle for the assessment, there was no requirement for equating work. Consequently, no pertinent evidence is provided for this element in the current context. However, it's worth highlighting that item calibration for the 2023 assessment employed the same well-established IRT scale that was put in place during the previous year.

Relevance: Evidence is highly relevant.

Completeness: Evidence is complete.

Overall Support: Existing evidence supports the assumption.

Assumption 1.5. Item and test scoring in 2023 were implemented accurately.

Element 1.5.1. Machine-scored items were scored accurately.

Machine-scored items undergo a key verification process as detailed in Chapter 6. This process involves a thorough review of all operational multiple-choice items before the scores are reported to confirm that the designated correct responses match the actual correct answers.

Relevance: Evidence is highly relevant.

Completeness: Evidence is complete.

Overall Support: Existing evidence strongly supports the assumption.

Assumption 1.6. MSAA Science scores correlate as expected with external indicators of student proficiency (i.e., concurrent and predictive evidence).

Element 1.6.1. MSAA Science scores correlate as expected with other measures of student proficiency.

Evidence: Peer reviewers acknowledge the challenge of acquiring additional evidence of student achievement that can be correlated with state alternate assessment scores, which they require for state grade-level assessments. As an alternative, peer reviewers do accept correlations that are internal to an alternate assessment as evidence in support of this assumption (D. Peasley, personal communication to S. Ferrara, October 21, 2019). The correlations between 2023 MSAA Science scale scores with the ELA and mathematics scores in grades 5, 8, and HS are listed in the table below.

Table 11-2. 2023 MSAA Science Scale Scores Correlations

	Correlations	
Grade	**ELA and Science**	**Math and Science**
5	0.86	0.82
8	0.84	0.81
HS	0.88	0.85

The strong positive values between the MSAA Science and the ELA and mathematics scale scores provide convergent validity evidence in the sense that they suggest that students' general academic and communicative capabilities are reflected strongly in both their MSAA Science, and their MSAA ELA and mathematics performances and scores.

Relevance: Evidence is moderately relevant.

Completeness: Evidence is limited.

Overall Support: Evidence provides limited support of the element. The lack of external assessment to correlate with MSAA Science is a possible challenge.

11.2 Primary Intended Score Uses

11.2.1 Primary Intended Score Use 1

Schools and districts use the MSAA and its results to (a) monitor trends in student performance and (b) design professional development for teachers.

Assumption 2.1. MSAA Science scores enable teachers and school, district, and state leaders to monitor trends in student proficiency.

2.1.1. MSAA Science scale scores for groups of students are adequately reliable and valid to enable school, district, and state leaders to monitor changes in means, standard deviations, and proficiency level percentages for classroom, school, district, and state groups.

2.1.2. MSAA Science 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.

2.1.3. The relationship between MSAA Science scores and external measures of student achievement and growth is as expected, compared to grade-level assessments and other measures of student achievement.

Individual score reliability in section 10.1 is comparable to industry standards, and aggregated scores (e.g., means) reliability is as high as or higher than individual scores. Additionally, section 10.2 discusses subgroup reliability, indicating strong reliability for some subgroups. However, caution is advised in interpreting subgroup score reliability due to potential issues arising from small subgroup sizes and restricted range.

The reliability and validity of MSAA Science scores and proficiency level categorizations for tracking student cohort performance can be assessed by examining the evidence provided in section 10.3. This section specifically emphasizes the consistency and accuracy in classifying performance levels.

One key aspect of this evaluation involves investigating the internal correlations between 2023 MSAA Science scores and scores in ELA and mathematics for students in grades 5, 8, and high school. These correlations serve as a measure of the relationship between MSAA Science scores and external indicators of student achievement. The findings reveal a significant correlation, ranging from moderate to strong. This outcome suggests that MSAA Science scores are a valuable tool for monitoring trends in student achievement across different content areas. In other words, the scores offer a reliable way to track how students are performing in science in relation to their performance in other subjects like ELA and mathematics.

Since the beginning of 2023, MSAA has partnered with Cognia to conduct a survey targeting district and school leaders, aiming to understand the utilization of MSAA scores in the context of monitoring trends in student proficiency. The initial phase of this validity study survey included participation from Arizona and Montana, with a total of 43 responses out of 100 surveys sent out, resulting in approximately a 30% response rate.

Key highlights from the pilot results include:

- Demographics: The majority of survey responses came from small school districts in rural areas, with approximately 70% of schools having fewer than 20 students participating in the MSAA. About 77% of respondents had extensive experience (more than 6 years) working with students with significant cognitive disabilities. The responses represented a range of grade levels (K–12), and special education directors and coordinators were the most prominent participants.
- Monitoring Trends in Student Data: All three types of reports (Individual Student Reports, school summary reports, and district summary reports) were used to monitor trends in student data. The Individual Student Report was the most frequently used, with a focus on performance levels and PLDs. In the district/school summary report, the mean scale score and the number of enrolled students received the most attention.

The pilot phase of the survey yielded limited information, primarily because it did not encompass the representative MSAA population. In the upcoming phase, which involves a larger group of MSAA partners, the operational survey is anticipated to offer a more comprehensive understanding of how teachers employ MSAA data in monitoring student progress.

Relevance: Evidence is highly relevant.

Completeness: Evidence is moderate to substantial.

Overall Support: Evidence moderately supports the element. Collecting additional evidence to correlate MSAA Science with external assessment may be needed.

Assumption 2.2. MSAA Science results are used to design professional development for teachers.

States offer guidance to local districts for developing teacher professional development, as exemplified by the NCSC's document titled "How to Teach the State Standards to Students Who Take Alternate Assessments" (accessible at https://www.azed.gov/sites/default/files/2016/12/How%20to%20Teach%20State%20Standards%20to%20Students%20Who%20Take%20Alternate%20Assessments%20030617%20Updated%20Links.pdf?id=5866dbe1aadebe085c4de5b4).

Furthermore, in the survey referenced in Assumption 2.1, there is additional evidence regarding the utilization of MSAA results in shaping professional development for educators from the pilot validity survey. It was found that only 40% of pilot survey participants reported offering professional development (PD) opportunities to teachers specifically focused on interpreting and applying MSAA scores. These PD sessions primarily served the purposes of aiding in the identification of Individualized Education Programs (IEPs) and the establishment of performance benchmarks.

Additionally, one-third of the respondents indicated that they conducted MSAA-related presentations, typically on an annual basis. These presentations were primarily targeted at teachers and school/district leaders.

Relevance: Evidence is highly relevant.

Completeness: Evidence is moderate to substantial.

Overall Support: Existing evidence moderately supports the assumption. It's noteworthy that the demographics of the pilot survey may not entirely reflect the broader MSAA population.

11.2.2 Primary Intended Score Use 2

The MSAA and its results are used to help teachers integrate MSAA scores and other information into their instructional planning.

Assumption 3.1. Teachers use the MSAA Science and its results to better integrate assessment with their instructional planning.

3.1.1. Teachers find the performance level descriptors and their students' performance levels useful for planning instruction, especially students in performance levels 1 and 2.

3.1.2. Teachers find their students' scale score information useful for planning instruction, especially students in Levels 1 and 2.

Special education teachers commonly rely on performance level descriptors (PLDs) to establish students' performance levels and shape instructional goals, particularly for those in performance levels 1 and 2. This practice is identified through annual compliance monitoring of Individualized Education Programs (IEPs) across states, exemplified by the Arizona Department of Education's requirement for measurable annual goals aligned with PLDs in IEPs. To gain a comprehensive understanding of the utility of MSAA scores and information for instructional planning, additional data, such as district/school leader surveys, are essential.

Relevance: Evidence is moderately relevant.

Completeness: Evidence is limited.

Overall Support: Existing evidence moderately supports the assumption. An example of additional evidence could be a survey of teachers to begin to understand the degree to which teachers find MSAA scores useful for planning instruction.

Assumption 3.2. Teachers use MSAA Science scores and other information for instructional planning.

Special education teachers often utilize MSAA scores and associated information for instructional planning, particularly in the context of establishing present levels of performance and developing goals, as indicated by annual compliance monitoring of Individualized Education Programs (IEPs) across states. Notably, the Arizona Department of Education mandates that IEPs incorporate measurable annual goals that align with performance level descriptors (PLDs). Additionally, teachers have access to MSAA teacher guides to assess student achievement and support instructional planning. However, while this evidence carries some relevance, its scope is limited. To obtain a comprehensive understanding of the degree to which teachers employ MSAA scores and associated information for planning instruction, further data collection methods such as teacher surveys are recommended.

Relevance: Evidence is moderately relevant.

Completeness: Evidence is limited.

Overall Support: Existing evidence moderately supports the assumption. An example of additional evidence could be a survey of teachers to begin to understand the degree to which teachers find MSAA scores useful for planning instruction.

11.2.3 Primary Intended Score Use 3

Parents use the MSAA and its results to get information about (a) what their child knows and can do, and (b) their child’s progress from year to year.

Assumption 4.1. Parents find MSAA Science scores and other information useful for understanding what their child knows and can do.

4.1.1. Parents understand and correctly interpret MSAA scores and other information to understand what their child knows and can do.

4.1.2. Parents use MSAA scores and other information appropriately to understand what their child knows and what their child can do and make decisions about their child’s education and learning needs.

MSAA provides assistance to parents in score interpretation and effectively utilizing MSAA scores and associated information to understand their child's achievements and educational needs. For instance, the Arizona Department of Education supplies Parent Overviews alongside each child's Individual Score Report, available in both English and Spanish. Similarly, the Maine Department of Education furnishes a Parent Overview of the MSAA Assessment System. However, the evidence's relevance is moderate, and while it supports the element to some extent, additional data, such as surveys of parents, are required to assess the extent to which parents correctly understand and use MSAA scores and related information for their child's educational decisions.

Relevance: Evidence is moderately relevant.

Completeness: Evidence is limited.

Overall Support: Existing evidence moderately supports the assumption. An example of additional evidence could be a survey of parents to begin to understand the degree to which parents correctly understand and interpret MSAA scores and other MSAA-based information to understand what their child knows and can do.

Assumption 4.2. Parents find MSAA Science scores and other information useful for understanding their child’s progress over time.

Element 4.2.1. Parents understand and correctly interpret MSAA Science scores and other information to understand their child’s progress over time.

Element 4.2.2. Parents use MSAA Science scores and other information appropriately to understand their child’s progress over time and make decisions about their child’s education and learning needs.

MSAA strives to assist parents in accurately interpreting and effectively utilizing MSAA scores and related information to comprehend their child's year-to-year progress and educational needs. For instance, the Arizona Department of Education provides Parent Overviews alongside each child's Individual Score Report, available in both English and Spanish. Similarly, the Maine Department of Education offers a Parent Overview of the MSAA Assessment System. However, the evidence's relevance is moderate, and while it partially supports the element, additional data, such as surveys of parents, are essential to assess the extent to which parents accurately understand and use MSAA scores and associated information for monitoring their child's progress and making informed decisions regarding their education and learning needs.

Relevance: Evidence is moderately relevant.

Completeness: Evidence is limited.

Overall Support: Existing evidence moderately supports the assumption. An example of additional evidence could be a survey of parents to begin to understand the degree to which parents correctly understand and interpret MSAA scores and other MSAA-based information to understand what their child knows and can do.

11.3 Conclusions

Because 2023 is the second year of the MSAA Science assessment, many assumptions and associated elements have limited or moderate evidence. However, both the quantity and quality of the evidence are expected to improve in the following years. The MSAA Psychometrics Subcommittee acknowledges areas where evidence may be weak or missing and has developed a research agenda to further develop evidence in those areas. These assumptions and elements form the validity arguments for MSAA scores, and their relevance, completeness, and overall support are summarized in Table 11.2 below.

Primary Score Intended Score Interpretation

MSAA Science scores provide reliable and valid information about important knowledge and skills in elementary, middle, and high school multidimensional science concepts that students with the most significant cognitive disabilities are attaining. All six assumptions supporting the intended score interpretation provide highly relevant evidence. Of these, five assumptions have complete evidence, while one has limited support. Moreover, five of the assumptions offer strong support of the intended score interpretation, with only one assumption having limited backing.

Intended Score Use 1

Schools and districts use the MSAA and its results to (a) monitor trends in student performance and (b) design professional development for teachers. Both of the assumptions backing intended score use 1 are well-supported, with highly relevant and complete evidence for each. Furthermore, both assumptions offer strongly supportive evidence overall for score use 1.

Intended Score Use 2

Teachers use the MSAA and its results to better integrate assessment with their instructional planning. Both of the assumptions supporting intended score use 2 exhibit moderately relevant and substantiated evidence, with both providing moderate to substantial support for score use 2.

Intended Score Use 3

Parents use the MSAA and its results to get information about (a) what their child knows and what their child can do and (b) their child's progress from year to year. Both of the assumptions backing intended score use 3 demonstrate moderately relevant and substantiated evidence. Moreover, they collectively contribute moderately to the overall support for score use 3.

Table 11-3. Status of Relevance of the Evidence for All Four SIUs, Assumptions, and Elements

Element	Relevance of the Evidence		Completeness of the Evidence to the Assumption			Overall Support to the Assumption		
	Highly Relevant	Moderately Relevant	Complete Evidence	Moderate to Substantial Evidence	Limited Evidence	Strongly Support	Moderately support	Limited Support
Primary Intended Score Interpretation MSAA scores provide reliable and valid information about important knowledge and skills in grade-level numeracy and literacy that students with the most significant cognitive disabilities are attaining.								
1.1 The content of the test represents the content of the standards (i.e., the Extended Performance Expectations).	X		X			X		
1.2 MSAA Science test items are construct relevant. The elements related to this assumption involve the skills and cognitive processes needed to respond to a specific item, and their alignment with those in the PLDs.	X		X			X		
1.3 Test administrations in MSAA states followed prescribed, standardized procedural requirements.	X		X			X		
1.4. Test scores on the MSAA Science provide reliable information about student performance and accurate classifications into performance levels.	X		X			X		
1.5 Item and test scoring were implemented accurately.	X		X			X		

continued

Element	Relevance of the Evidence		Completeness of the Evidence to the Assumption			Overall Support to the Assumption		
	Highly Relevant	Moderately Relevant	Complete Evidence	Moderate to Substantial Evidence	Limited Evidence	Strongly Support	Moderately support	Limited Support
1.6 MSAA Science scores correlate as expected with external indicators of student proficiency (i.e., concurrent evidence).	X				X			X
Primary Intended Score Use 1 Schools and districts use the MSAA and its results to (a) monitor trends in school performance and (b) design professional development for teachers.								
2.1 Schools and districts use the MSAA Science and its results to monitor trends in school performance.	X			X			X	
2.2 MSAA Science results are used to design professional development for teachers.	X			X			X	
Primary Intended Score Use 2 The MSAA and its results are used to help teachers integrate MSAA scores and other information with their instructional planning.								
3.1 Teachers use the MSAA Science and its results to better integrate assessment with their instructional planning.		X			X		X	
3.2 Teachers use MSAA Science scores and other information for instructional planning.		X			X		X	
Primary Intended Score Use 3 Parents understand and interpret correctly MSAA scores and other information to understand what their child knows and can do.								
4.1. Parents find MSAA Science scores and other information useful for understanding what their child knows and can do.		X			X		X	
4.2. Parents find MSAA Science scores and other information useful for understanding their child's progress over time.		X			X		X	

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory.* Belmont, CA: Wadsworth, Inc.

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.

Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

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.

Cameto, R., Haertel, G., Morrison, K., & Russell, M. (2010). *Synergistic use of evidence-centered design and universal design for learning for improved assessment design.* Menlo Park, CA: SRI International.

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

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

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

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Ferrara, S., Lai, E., Reilly, A., & Nichols, P. D. (2016). Principled approaches to assessment design, development, and implementation: Cognition in score interpretation and use. In A. A. Rupp and J. P. Leighton (Eds.), *The handbook on cognition and assessment: Frameworks, methodologies, & applications* (pp. 41–74). Malden, MA: Wiley.

Ferrara, S., & Qunbar, S. (2022). Validity Arguments for AI-Based Automated Scores: Essay Scoring as an Illustration. *Journal of Educational Measurement*, *59*(3), 288-313.

Flowers, C., Wakeman, S., Browder, D., & Karvonen, M. (2007). Links for academic learning: An alignment protocol for alternate assessments based on alternate achievement standards. Charlotte, NC: University of North Carolina at Charlotte. Retrieved from: http://www.naacpartners.org/LAL/documents/NAAC_AlignmentManualVer8_3.pdf.

Forte, E. (2013a). *Evaluating alignment for assessments developed using evidence-centered design*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA.

Forte, E. (2013b). *The next generation of alignment.* Paper presented at the National Conference on Student Assessment, Washington, DC.

Gulliksen, H. (1950). *Theory of Mental Tests*. New York, NY: John Wiley and Sons.

GSS Lead States. 2013. Next Generation Science Standards: For States, By States. Washington, DC: The National Academies Press

Hagge, S., & Keller, R. (2009). *Equating mixed-format tests: Examining the impact of intra-individual consistency in the IRT framework*. Presented at the American Educational Research Association, San Diego, CA.

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

Hambleton, R. K., Swaminathan, H., & Rogers, J. H. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

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, DC: National Council on Measurement in Education.

Johnstone, C. J., Altman, J., & Thurlow, M. (2006). *A state guide to the development of universally designed assessments*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

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

Keller, L., Keller, R., & Parker, P. (2008). *The effect of shifting content on the accuracy of equating.* Presented at the American Educational Research Association, New York, NY.

Keller, R., Kim, W., Nering, M., & Keller, L. (2007). *What breaks the equating? A preliminary investigation into threats to a five-year equating chain.* Presented at the American Educational Research Association, Chicago, IL.

Kleinert, H. L., Browder, D. M., & Towles-Reeves, E. A. (2009). Models of cognition for students with significant cognitive disabilities: Implications for assessment. *Review of Educational Research, 79,* 301–326.

Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). New York, NY: Routledge.

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York, NY: Springer-Verlag.

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. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Measured Progress Department of Psychometrics and Research (2015). *National Center and State Collaborative (NCSC) standard setting report*. Unpublished report.[2]

Michaelides, P. M. (2003). *Sensitivity of IRT equating to the behavior of test equating item.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Muraki, E., & Bock, R. D. (2003). PARSCALE 4.1 [Computer software]. Lincolnwood, IL: Scientific Software International.

National Research Council. 2012. A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC: The National Academies Press. https://nap.nationalacademies.org/catalog/13165/a-framework-for-k-12-science-education-practices-crosscutting-concepts.

[2]*See Chapter 7:* http://ncscpartners.org/Media/Default/PDFs/Resources/NCSC15_NCSC_TechnicalManualNarrative.pdf

Parker, P., Keller, R., & von Davier, A. (2009). *The Examination of four equating methods: The effects of reclassifying students into performance categories and the population sensitivity assumption*. Presented at the National Council on Measurement in Education, San Diego, CA.

Persuasiveness. (2019). In *Dictionary.com* Retrieved from https://www.dictionary.com/browse/plausible

Plausibility. (2019). In *Dictionary.com* Retrieved from https://www.dictionary.com/browse/plausible

Relevance. (2019). In *Dictionary.com* Retrieved from https://www.dictionary.com/browse/relevance

Rudner, L. M. (2001). Computing the expected proportions of misclassified examinees. *Practical Research & Evaluation, 7* (14). Retrieved from http://PARE.online.net/getvn.asp?v=7&n=14

Rudner, L. M. (2005). Expected classification accuracy. *Practical Research & Evaluation, 10* (13). Retrieved from http://pareonline.net/pdf/v10n13.pdf

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.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*, 201–210.

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

Stout, W. 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.

Wakeman, S., Lee, A., Browder, B. (2012, August 23). *National Center and State Collaborative approach to Content for Students with Significant Disabilities*. National Center and State Collaborative. http://ncscpartners.org/Media/Default/PDFs/Resources/Wakeman-etal-CEC2012-NCSC-Approach-to-Content-SWSCD.pdf

Webb, N. L. (2005, November). *Alignment, depth of knowledge, and change*. Paper presented at the annual meeting of the Florida Educational Research Association, Miami, Florida.

Yan, D., von Davier, A. A., & Lewis, C. (Eds.). (2014). Computerized multistage testing: Theory and applications. Boca Raton: CRC 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
ACCOMMODATION FREQUENCIES

Table A-1. Accommodation Frequencies

	Grades		
Accommodations	**5**	**8**	**11**
LCI_Vision[1]	63	68	69
SAR_Assistive_Response_After[2]	140	141	120
SAR_No_Accomm_Needed_After[3]	214	208	207
SAR_Paper_Version_After[4]	1	0	1
SAR_Scribe_After[5]	61	61	28
SAR_Sign_Interpretation_After[6]	196	203	111

1: LCI_Vision - Input could occur through alternate keyboards, eye-gaze, switch devices, speech-to-text, and other similar input devices. Students are also expected to access text using AT devices (e.g., screen readers), but refreshable Braille display is not supported for presentation of text-based content for the first operational year.
2: SAR_Assistive_Response_After - Assistive Technology (AT) for viewing, responding, or interacting with test items.
3: SAR_No_Accomm_Needed_After - No accommodations needed.
4: SAR_Paper_Version_After - Paper version of item/s.
5: SAR_Scribe_After - A scribe will enter in the MSAA Online Assessment System the student-indicated answer to a selected-response item. For the constructed-response writing item, the scribe will record the student's response to the writing prompt on the response templates in the MSAA Online Assessment System.
6: SAR_Sign_Interpretation_After - TA may communicate passages, items and response options using sign language to student.

Table A-2. Accommodation Summary

	Number of Students Tested	
Grade	**With Accommodations**	**Without Accommodations**
5	545	497
8	555	444
11	458	453

APPENDIX B

PANELISTS AND COMMITTEE MEMBERS

Table B-1. MSAA 2016 Test Blueprints, Design, and Draft EPEs Review Meeting Final Panelists

Science Content	
Name	**Expertise**
Susan Izard	SPED
Mariann Bell	SPED
Karen Travers-Lynch	Science Content
Paul Ritchie	Science Content

Table B-2. MSAA 2017 Item Specifications and Item Prototype Review Meeting Final Panelists

Science Content	
Name	**Expertise**
Yvonne Fields	SPED
Ashley McGrath	Science Content
Christina Booth	Science Content
Sam Shaw	Science Content
Mariann Bell	SPED
Betsy Rogers	SPED
Karen Travers-Lynch	Science Content
Paul Ritchie	Science Content

Table B-3. MSAA 2018 Item Content and Bias-Sensitivity Review Meeting Final Panelists by Grade and Content Area

Science Content Grades 5 & 8	
Name	**Expertise**
Melissa Lamont	SPED
Benjamin Altsher	SPED
Haley Johnson	SPED
Michelle DeBlois	Science Content
Michelle McCarthy	Science Content
Science Content High School	
Name	**Expertise**
Jim Flatten	Science Content
Thyra Galli	Science Content
Mariann Bell	SPED
Sue Nay	SPED
Karen Travers-Lynch	Science Content
Thyra Galli	Science Content

Table B-4. MSAA 2021 Item Content and Bias-Sensitivity Review Meeting Final Panelists by Grade and Content Area

Science Content Grade 5	
Name	**Expertise**
Vince McGroary	SPED
Kevin Fillion	SPED
Thyra Galli	Science Content
Andrea Kuegel	Science Content
Mckayla Hogan	Science Content
Krista Rowley	SPED & Science Content
Jodi Bossio Smith	SPED & Science Content
Bethany Spangenberg	SPED & Science Content
Science Content Grade 8	
Name	**Expertise**
L'Aura Routsong	SPED
Matt Arnold	SPED
Josh Weller	SPED
Brian Pixley	Science Content
Eugene Chan	Science Content
Shari Templeton	Science Content
Science Content Grade 11	
Name	**Expertise**
Robin Davis	SPED
Sarah Juhlin	Science Content
Andrea McClure	Science Content
Steve Ruback	Science Content
Bethany Spangenberg	SPED & Science Content

Table B-5. MSAA 2022 Item Content and Bias-Sensitivity Review Meeting Final Panelists by Grade and Content Area

Science Content Grades 5	
Name	**Expertise**
Lacey Todd	SPED & Science Content
Benjamin Altsher	SPED
Jeryline Aves	SPED
Janice Knox	Science Content
Tristan John Antonio David	Science Content
Jodi Bossio Smith	SPED & Science Content
Science Content Grades 8	
Name	**Expertise**
Andrea Mercado	Science Content
Agosto Jerusalem	SPED & Science Content
Don Griffin	SPED
Rhonda Bowman	SPED & Science Content
Katie Neate	SPED
Thyra Galli	Science Content
Science Content Grade 11	
Name	**Expertise**
Julie LaRosa	Science Content
Ellen Anfone	SPED
Lori Furr	Science Content
Benjamin Altsher	SPED
Kristen Nash	SPED

Table B-6. MSAA 2023 Item Content and Bias-Sensitivity Review Meeting Final Panelists by Grade and Content Area

Science Content Grades 5	
Name	**Expertise**
Cassandra Laapui	SPED & Science Content
Seth Blewitt	Science Content
Gerry Gonzales	Science Content
Mark Miskell	Science Content
Spring Cooke	Science Content
Christine C Hernandez	SPED
Helene S A Cruz	SPED
Denise Wilmot	SPED
Science Content Grades 8	
Name	**Expertise**
Senerita Kaleopa	SPED & Science Content
Anntonelli Pola	SPED & Science Content
Kristine Ladra	SPED
Denise Wilmot	SPED & Science Content
Bailey Brewster	SPED
Science Content Grade 11	
Name	**Expertise**
Tagiilima Uikirifi	SPED & Science Content
Roque Indalecio	Science Content
Anita Franklin	SPED
Mary-Alice Corliss	Science Content

Table B-7. MSAA 2023 Technical Advisory Committee Members

Name	Organization	Expertise
Derek Briggs	University of Colorado	• Assessment • Growth • Psychometrics
Chris Domaleski	Center for Assessment	• Accountability Systems • Psychometrics
Rachel Quenemoen	National Center on Educational Outcomes	• Students with Significant Cognitive Difficulties • NCSC Awareness
Mike Russell	Boston College	• Technology • Accessibility
Martha Thurlow	University of Minnesota/NCEO	• Special Education • Accessibility

APPENDIX C
TEST PARTICIPATION

Table C.1 Summary of Tested Students by Demographic Category

Description	Tested # Complete	Tested # No Observable Mode of Communication[1]	Total Tested	Total Percent
All Students	2,795	157	2,952	100
Female	997	58	1,055	36
Male	1,788	99	1,887	64
Gender Undefined	10	0	10	0
Hispanic or Latino	1,085	53	1,138	39
American Indian or Alaska Native	157	12	169	6
Asian	73	7	80	3
Black or African American	212	10	222	8
Native Hawaiian or Pacific Islander	53	2	55	2
White (non-Hispanic)	1,056	62	1,118	38
Two or More Races (non-Hispanic)	110	9	119	4
No Primary race/Ethnicity Undefined	49	2	51	2
Currently receiving LEP services	209	10	219	7
Not receiving LEP services	432	33	465	16
LEP: All Other Students	2,154	114	2,268	77
Economically Disadvantaged Students	243	15	258	9
Non-economically Disadvantaged Students	203	21	224	8
SES: All Other Students	2,349	121	2,470	84
Migrant	1	0	1	0
Non- migrant	403	36	439	15
Undefined Migrant Status	2,391	121	2,512	85
Augmentative Communication	596	57	653	22
No Augmentative Communication	2,187	98	2,285	77
Undefined Augmentative Communications	12	2	14	0
Hearing Loss	91	27	118	4
Within Normal Limits	2,704	129	2,833	96
Undefined Hearing Loss	0	1	1	0
Visual Impairment	131	69	200	7
Within Normal Limits	2,653	86	2,739	93
Undefined Visual Impairment	11	2	13	0
Sensory Stimuli Response	203	111	314	11
Follow Directions	2,592	45	2,637	89
Undefined Receptive Language	0	1	1	0
Special School	256	25	281	10
Regular School Self-contained	1,716	106	1,822	62
Regular School Resource Room	469	15	484	16
Regular School Primarily Self-contained	210	5	215	7
Regular School General Education	144	5	149	5
Undefined Classroom Setting	0	1	1	0
Student Communicates Primarily Through Cries	177	91	268	9
Uses Intentional Communication	686	42	728	25
Uses Symbolic Language	1,932	23	1,955	66
Undefined Expressive Communication	0	1	1	0

[1] No Observable Mode of Communication indicates that the students' test was closed because they had no visible means of communication.

Table C-2. Students Tested by Subgroup

Description	Total Tested	Invalidated	Did Not Test
Science	2,952	34	261

APPENDIX—D

DETAILED CONTENT RATIONALE FOR PERFORMANCE EXPECTATION SELECTION FOR ALL GRADES

AND

ELEMENTARY GRADE-LEVEL EXTENDED PERFORMANCE EXPECTATIONS EXAMPLE

The Cognia development team followed a principled assessment design process and utilized the published national resources of the *Framework* and NGSS. As outlined in Ferrara, Lai, Reilly, and Nichols (2016), "principled approaches provide concepts, procedures, and tools to guide assessment design, development, and implementation decisions" (pg. 3). The test design and test blueprint provided the guardrails for the selection of the Performance Expectations (PEs) and creation of the Extended Performance Expectations (EPEs).

The table below shows the content blueprint for the operational test for each grade band. The test blueprint for each grade in Appendix F incorporates the overall content distributions used for the development of the operational tests.

Table D-1. Blueprint for Distribution of Science Content by Grade Level

Science Content Category	Grade 5	Grade 8	HS
Physical Sciences	~40%	30-40%	30-40%
Life Sciences	~30%	30-40%	30-40%
Earth and Space Sciences	~30%	~30%	~30%

The selected standards assessable within each category, and the rationale for their selection, are further explained in the next part of this appendix. Following detail about the PE selection rationale an example of the EPEs along with an explanation of the major layout components is provided.

For the grade 5 test, PEs from grades 3, 4, and 5 are included; the progression of standards in those grades is such that to provide a solid representation of the core ideas and understandings students need to progress from elementary school to middle school, PEs needed to be selected across grade bands. The table below shows the collection of PEs chosen to be assessed on the grade 5 test.

Table D-2. Selected Performance Expectations for Grade 5 Test

Performance Expectation (PE)	DCI	SEP	CCC
5-PS1-2. Measure and graph quantities to provide evidence that regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is conserved.	PS1.A Structure and Properties of Matter PS1.B Chemical Reactions	Using Mathematics and Computational Thinking	Scale, Proportion, and Quantity
3-PS2-2. Make observations and/or measurements of an object's motion to provide evidence that a pattern can be used to predict future motion.	PS2.A Forces and Motion	Planning and Carrying Out Investigations	Patterns
5-PS2-1. Support an argument that the gravitational force exerted by Earth on objects is directed down.	PS2.B Types of Interactions	Engaging in Argument from Evidence	Cause and Effect
4-PS3-4. Apply scientific ideas to design, test, and refine a device that converts energy from one form to another. *	PS3.B Conservation of Energy and Energy Transfer PS3.D Energy in Chemical Processes and Everyday Life ETS1.A Defining and Delimiting an Engineering Problem	Designing Solutions	Energy and Matter
5-PS3-1. Use models to describe that energy in animals' food (used for body repair, growth, and motion, and to maintain body warmth) was once energy from the sun.[1]	PS3.D Energy in Chemical Processes and Everyday Life	Developing and Using Models	Energy and Matter

continued

Performance Expectation (PE)	DCI	SEP	CCC
4-LS1-1. Construct an argument that plants, and animals have internal and external structures that function to support survival, growth, behavior, and reproduction.	LS1.A Structure and Function	Engaging in Argument from Evidence	Systems and System Models
3-LS3-1. Analyze and interpret data to provide evidence that plants and animals have traits inherited from parents and that variation of these traits exists in a group of similar organisms.	LS3.A Inheritance of Traits LS3.B Variation of Traits	Analyzing and Interpreting Data	Patterns
3-LS4-1. Analyze and interpret data from fossils to provide evidence of the organisms and environments in which they lived long ago.	LS4.A Evidence of Common Ancestry and Diversity	Analyzing and Interpreting Data	Scale, Proportion, and Quantity
5-ESS1-2. Represent data in graphical displays to reveal patterns of daily changes in the length and direction of shadows, day and night, and the seasonal appearance of some stars in the night sky.	ESS1.B Earth and the Solar System	Analyzing and Interpreting Data	Patterns
3-ESS2-1. Represent data in tables and graphical displays to describe typical weather conditions expected during a particular season.	ESS2.D Weather and Climate	Analyzing and Interpreting Data	Patterns
5-ESS2-1. Develop a model using an example to describe ways in which the geosphere, biosphere, hydrosphere, and/or atmosphere interact.	ESS2.A Earth Materials and Systems	Developing and Using Models	Systems and System Models
5-ESS3-1. Obtain and combine information about ways individual communities use science ideas to protect the Earth's resources and environment.	ESS3.C Human Impacts on Earth Systems	Obtaining, Evaluating, and Communicating Information	Systems and System Models

**PEs marked with an asterisk incorporate Engineering constructs.*
[1]This PE crosses Physical Sciences and Life Sciences. It will be classified in Life Sciences in fulfilling the blueprint distribution.

For the grade 5 test, PEs were chosen from grades 3-5 to generate the best representation of broad, fundamental principles for the elementary grade assessment. Because the NGSS spread science topics out across grades in the 3-5 grade band, there are some key content ideas for forces and motion, as well as heredity, biological evolution, and weather, that only appear in standards at grade 3. Although the test is targeted for administration to students in grade 5, four grade 3 PEs were included to ensure focus on all foundational areas that students would need exposure to, to prepare for middle school expectations:

- 3-PS2-2 focuses on basic patterns of motion, as a foundation of the cause-and-effect exploration of forces and motion. This PE also provides an opportunity to expose students to the various types of forces, from physical contact forces to gravity and magnetism, linking to another motion/forces PE within elementary and to other motion/forces PEs in later grades.
- 3-LS3-1 introduces the fundamental principle of inheritance of traits (traits pass from parents to offspring) as well as the idea of variation, which are both cornerstones of the study of genetics and biological evolution.
- 3-LS4-1 provides an accessible foundation for thinking about evidence of organisms' fit to the environment, and changes in organisms and environments over time.
- 3-ESS2-1 focuses on the most foundational understandings of weather, which are then extended in other elementary PEs and in later grades in studying interactions of Earth's systems, geoscience processes changing Earth's surface, water cycling through Earth's systems, and the larger concept of climate.

It should also be noted that while the chosen PEs may seem to lean more heavily to Physical Sciences than Life Sciences (5 PEs coded to Physical Sciences and only 3 PEs coded to Life Sciences), PE 5-PS3-1 is a “crossover” PE that connects the physical science concept of energy in everyday life with the life science concept of matter and energy flow. Although 5-PS3-1 has a physical science coding, it would typically be taught within an ecology unit (and is therefore classified as a Life Science PE in the test blueprint).

Additionally, it may be noted that there are no PEs in the elementary grade test for Physical Sciences DCI PS4, Waves and Their Application in Technologies for Information Transfer. The concept of waves is abstract and is therefore viewed as more appropriate to address in the grade 8 test than in this grade band for students with significant cognitive disabilities. Likewise in Life Sciences, although no performance expectation is explicitly aligned to DCI LS2 (Ecosystems: Interactions, Energy, and Dynamics), PE 5-PS3-1 actually overlaps heavily with these concepts. All other DCIs are represented in the elementary grade test. The table below, shows the DCIs included in the grade 5 test.

Table D-3. Disciplinary Core Idea Summary for Grade 5 Test

Physical Sciences													
PS1.A	PS1.B	PS1.C	PS2.A	PS2.B	PS2.C	PS3.A	PS3.B	PS3.C	PS3.D	PS4.A	PS4.B	PS4.C	
X	X		X	X			X		X				
Life Sciences													
LS1.A	LS1.B	LS1.C	LS1.D	LS2.A	LS2.B	LS2.C	LS2.D	LS3.A	LS3.B	LS4.A	LS4.B	LS4.C	LS4.D
X								X	X	X			
Earth and Spaces Sciences													
ESS1.A	ESS1.B	ESS1.C	ESS2.A	ESS2.B	ESS2.C	ESS2.D	ESS2.E	ESS3.A	ESS3.B	ESS3.C	ESS3.D		
	X		X			X				X			
Engineering Design													
ETS1.A	ETS1.B	ETS1.C											
X													

The representation of SEPs and CCCs across the selected PEs was also reviewed to ensure most, if not all, SEPs and CCCs were included for each grade test. Likewise, one or two engineering-aligned PEs were included in the selected PEs for each grade test, as engineering constructs are included in the *Framework* as both SEPs and DCIs. As can be seen in Table D-2, in the grade 5 test, the selected PEs incorporate seven of the eight SEPs and five of the seven CCCs (with the other two CCCs not actually included in the elementary grade band in NGSS standards). There is also one engineering-aligned PE included, 4-PS3-4.

For the grade 8 test, four PEs were selected per content domain. The PEs were chosen from the middle school grade band to generate the best representation of important principles that bridge the elementary and high school grades, and that are accessible and relevant for this student population. The table below shows the collection of PEs chosen to be assessed on the grade 8 test.

Table D-4. Selected Performance Expectations for Grade 8 Test

Performance Expectation (PE)	DCI	SEP	CCC
MS-PS1-2. Analyze and interpret data on the properties of substances before and after the substances interact to determine if a chemical reaction has occurred.	PS1.A Structure and Properties of Matter PS1.B Chemical Reactions	Analyzing and Interpreting Data	Patterns
MS-PS2-2. Plan an investigation to provide evidence that the change in an object's motion depends on the sum of the forces on the object and the mass of the object.	PS2.A Forces and Motion	Planning and Carrying Out Investigations	Stability and Change
MS-PS3-5. Construct, use, and present arguments to support the claim that when the kinetic energy of an object changes, energy is transferred to or from the object.	PS3.B Conservation of Energy and Energy Transfer	Engaging in Argument from Evidence	Energy and Matter
MS-PS4-2. Develop and use a model to describe that waves are reflected, absorbed, or transmitted through various materials.	PS4.A Wave Properties PS4.B Electromagnetic Radiation	Developing and Using Models	Structure and Function
MS-LS1-3. Use argument supported by evidence for how the body is a system of interacting sub-systems composed of groups of cells.	LS1.A Structure and Function	Engaging in Argument from Evidence	Systems and System Models
MS-LS1-5. Construct a scientific explanation based on evidence for how environmental and genetic factors influence the growth of organisms.	LS1.B Growth and Development of Organisms	Constructing Explanations	Cause and Effect
MS-LS2-1. Analyze and interpret data to provide evidence for the effects of resource availability on organisms and populations of organisms in an ecosystem.	LS2.A Interdependent Relationships in Ecosystems	Analyzing and Interpreting Data	Cause and Effect
MS-LS2-3. Develop a model to describe the cycling of matter and flow of energy among living and non-living parts of an ecosystem.	LS2.B Cycle of Matter and Energy Transfer in Ecosystems	Developing and Using Models	Energy and Matter
MS-ESS1-1. Develop and use a model of the Earth-sun-moon system to describe the cyclic patterns of lunar phases, eclipses of the sun and moon, and seasons.	ESS1.A The Universe and Its Stars ESS1.B Earth and the Solar System	Developing and Using Models	Patterns
MS-ESS2-2. Construct an explanation based on evidence for how geoscience processes have changed Earth's surface at varying time and spatial scales.	ESS2.A Earth Materials and Systems ESS2.C The Role of Water in Earth's Surface Processes	Constructing Explanations	Scale, Proportion, and Quantity
MS-ESS2-4. Develop a model to describe the cycling of water through Earth's systems driven by energy from the sun and the force of gravity.	ESS2.C The Role of Water in Earth's Surface Processes	Developing and Using Models	Energy and Matter
MS-ESS3-3. Apply scientific principles to design a method for monitoring and minimizing a human impact on the environment.*	ESS3.C Human Impacts on Earth Systems ETS1.A Defining and Delimiting an Engineering Problem	Designing Solutions	Cause and Effect

Across the content domains, choices for representation of important principles that bridge the elementary and high school grades, and that are accessible and relevant for this student population, include the following:

- In Physical Sciences, DCI PS4 (Waves and Their Application in Technologies for Information Transfer) – which was not assessed in the grade 5 test – is now addressed with an accessible, grade-appropriate standard (MS-PS4-2) focused on reflection, absorption, and transmission of waves through materials.
- In Life Sciences, PEs for DCIs LS3 and LS4 are not included. A strong focus on body systems and growth (DCI LS1) and ecological principles (DCI LS2) is viewed as more accessible and relevant for students in the middle grades than emphases for heredity (LS3) and biological

evolution (LS4) in this grade band. Many of the middle-grade PEs for LS3 and LS4 are conceptually difficult and abstract. Rather than deconstruct these PEs down to a point that provides less-than-optimal learning and development opportunity for students, these PEs have not been prioritized at the middle school level. The LS3 and LS4 core ideas are being addressed grade-appropriately in the high school tests.

As shown in table below, concepts for all DCIs are included in the grade 8 test with the exception of LS3 and LS4, as explained above.

Table D-5. Disciplinary Core Idea Summary for Grade 8 Test

Physical Sciences													
PS1.A	PS1.B	PS1.C	PS2.A	PS2.B	PS2.C	PS3.A	PS3.B	PS3.C	PS3.D	PS4.A	PS4.B	PS4.C	
X	X		X				X			X	X		
Life Sciences													
LS1.A	LS1.B	LS1.C	LS1.D	LS2.A	LS2.B	LS2.C	LS2.D	LS3.A	LS3.B	LS4.A	LS4.B	LS4.C	LS4.D
X	X			X	X								
Earth and Spaces Sciences													
ESS1.A	ESS1.B	ESS1.C	ESS2.A	ESS2.B	ESS2.C	ESS2.D	ESS2.E	ESS3.A	ESS3.B	ESS3.C	ESS3.D		
X	X		X		X					X			
Engineering Design													
ETS1.A	ETS1.B	ETS1.C											
X													

The representation of SEPs and CCCs across the selected PEs was also reviewed to ensure most, if not all, SEPs and CCCs were included for each grade. Likewise, one or two engineering-aligned PEs were included in the selected PEs for each grade test, as engineering constructs are included in the *Framework* as both SEPs and DCIs. As can be seen in Table D-4, in the grade 8 test, the selected PEs incorporate five of the eight SEPs and all seven of the CCCs. For the SEPs, there was a trade-off of optimizing coverage of key content ideas and coverage of Practices, and key content coverage was prioritized. Using Mathematics and Computational Thinking is not represented for the highest access point for grade 8, as the focal skills of that Practice that would be most appropriate for students with significant cognitive disabilities heavily overlap those of the SEP Analyzing and Interpreting Data. The SEP Asking Questions and Defining Problems and the SEP Obtaining, Evaluating, and Communicating Information do not appear very frequently in the middle school PEs compared to the other PEs. To include engineering, there is also one engineering-aligned PE included, MS-ESS3-3.

For the high school test, four PEs were selected per content domain. The PEs were chosen from the high school grade band to generate the best representation of important culminating understandings that are accessible and relevant for this student population. The table below shows the collection of PEs chosen to be assessed on the high school test.

Table D-6. Selected Performance Expectations for High School Test

Performance Expectation (PE)	DCI	SEP	CCC
HS-PS1-2. Construct and revise an explanation for the outcome of a simple chemical reaction based on the outermost electron states of atoms, trends in the periodic table, and knowledge of the patterns of chemical properties.	PS1.A Structure and Properties of Matter PS1.B Chemical Reactions	Constructing Explanations	Patterns
HS-PS2-3. Apply scientific and engineering ideas to design, evaluate, and refine a device that minimizes the force on a macroscopic object during a collision.*	PS2.A Forces and Motion ETS1.A Defining and Delimiting an Engineering Problem ETS1.C Optimizing the Design Solution	Designing Solutions	Cause and Effect
HS-PS2-5. Plan and conduct an investigation to provide evidence that an electrical current can produce a magnetic field and that a changing magnetic field can produce an electrical current.	PS2.B Types of Interactions PS3.A Definitions of Energy	Planning and Carrying Out Investigations	Cause and Effect
HS-PS3-2. Develop and use models to illustrate that energy at the macroscopic scale can be accounted for as a combination of energy associated with the motions of particles (objects) and energy associated with the relative positions of particles (objects).	PS3.A Definitions of Energy	Developing and Using Models	Energy and Matter
HS-LS2-2. Use mathematical representations to support and revise explanations based on evidence about factors affecting biodiversity and populations in ecosystems of different scales.	LS2.A Interdependent Relationships in Ecosystems	Using Mathematics and Computational Thinking	Scale, Proportion, and Quantity
HS-LS3-1. Ask questions to clarify relationships about the role of DNA and chromosomes in coding the instructions for characteristic traits passed from parents to offspring.	LS3.A Inheritance of Traits LS1.A Structure and Function	Asking Questions	Cause and Effect
HS-LS4-1. Communicate scientific information that common ancestry and biological evolution are supported by multiple lines of empirical evidence.	LS4.A Evidence of Common Ancestry and Diversity	Obtaining, Evaluating, and Communicating Information	Patterns
HS-LS4-3. Apply concepts of statistics and probability to support explanations that organisms with an advantageous heritable trait tend to increase in proportion to organisms lacking this trait.	LS4.B Natural Selection LS4.C Adaptation	Analyzing and Interpreting Data	Patterns
HS-ESS1-6. Apply scientific reasoning and evidence from ancient Earth materials, meteorites, and other planetary surfaces to construct an account of Earth's formation and early history.	ESS1.C The History of Planet Earth	Constructing Explanations	Stability and Change
HS-ESS2-4. Use a model to describe how variations in the flow of energy into and out of Earth's systems result in changes in climate.	ESS1.B Earth and the Solar System ESS2.A Earth Materials and Systems ESS2.D Weather and Climate	Developing and Using Models	Cause and Effect
HS-ESS2-5. Plan and conduct an investigation of the properties of water and its effects on Earth materials and surface processes.	ESS2.C The Role of Water in Earth's Surface Processes	Planning and Carrying Out Investigations	Structure and Function
HS-ESS3-4. Evaluate or refine a technological solution that reduces impacts of human activities on natural systems.*	ESS3.C Human Impacts on Earth Systems ETS1.B Developing Possible Solutions	Designing Solutions	Stability and Change

In Physical Sciences, DCI PS4 (Waves and Their Application in Technologies for Information Transfer) is addressed with many PEs that are extended topics and very abstract, and thus was not prioritized for this population of students compared to other content within the Physical Sciences. Foundational concepts for waves are covered in the middle grade band instead. Concepts for all other DCIs in each domain are included in the high school test, as shown in the table below.

Table D-7. Disciplinary Core Idea Summary for High School Test

Physical Sciences													
PS1.A	PS1.B	PS1.C	PS2.A	PS2.B	PS2.C	PS3.A	PS3.B	PS3.C	PS3.D	PS4.A	PS4.B	PS4.C	
X	X		X	X		X							
Life Sciences													
LS1.A	LS1.B	LS1.C	LS1.D	LS2.A	LS2.B	LS2.C	LS2.D	LS3.A	LS3.B	LS4.A	LS4.B	LS4.C	LS4.D
X				X				X		X	X	X	
Earth and Spaces Sciences													
ESS1.A	ESS1.B	ESS1.C	ESS2.A	ESS2.B	ESS2.C	ESS2.D	ESS2.E	ESS3.A	ESS3.B	ESS3.C	ESS3.D		
	X	X	X		X	X				X			
Engineering Design													
ETS1.A	ETS1.B	ETS1.C											
X	X	X											

The representation of SEPs and CCCs across the selected PEs was also reviewed to ensure most, if not all, SEPs and CCCs were included for each grade. Likewise, one or two engineering-aligned PEs were included in the selected PEs for each grade test, as engineering constructs are included in the *Framework* as both SEPs and DCIs. As can be seen in Table D-6, in the high school test, the selected PEs incorporate seven of the eight SEPs and six of the seven CCCs. The only SEP that is not represented for the highest access point for high school is Engaging in Argument from Evidence. This Practice is included in less than 1/8 of the PEs in the high school grade band, and key content coverage was prioritized over Practices in this case. The only CCC not included is Systems and System Models; however, because two of the PEs (HS-PS3-2 and HS-ESS2-4) integrate the Practice of Developing and Using Models, students will be applying systems thinking in those cases as well. In high school, there are also 2 engineering-aligned PEs included, HS-PS2-3 and HS-ESS3-4.

As PEs were selected and finalized for each grade band, the progression of DCIs across all grades was checked to help validate the appropriateness of the collection of PEs chosen for assessment on each grade's test. The tables below show an example of the final prioritized PEs and associated DCIs for the Physical Sciences across all three grades, followed by the same information for Life Science and Earth and Space Sciences.

Table D-8. Selected Performance Expectations for Physical Science Across Grades 5, 8, and HS

Grade	Performance Expectation (PE)	DCI
5	5-PS1-2. Measure and graph quantities to provide evidence that regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is conserved.	PS1.A PS1.B
	3-PS2-2. Make observations and/or measurements of an object's motion to provide evidence that a pattern can be used to predict future motion.	PS2.A
	5-PS2-1. Support an argument that the gravitational force exerted by Earth on objects is directed down.	PS2.B
	4-PS3-4. Apply scientific ideas to design, test, and refine a device that converts energy from one form to another. *	PS3.B PS3.D ETS1.A
8	MS-PS1-2. Analyze and interpret data on the properties of substances before and after the substances interact to determine if a chemical reaction has occurred.	PS1.A PS1.B
	MS-PS2-2. Plan an investigation to provide evidence that the change in an object's motion depends on the sum of the forces on the object and the mass of the object.	PS2.A
	MS-PS3-5. Construct, use, and present arguments to support the claim that when the kinetic energy of an object changes, energy is transferred to or from the object.	PS3.B
	MS-PS4-2. Develop and use a model to describe that waves are reflected, absorbed, or transmitted through various materials.	PS4.A PS4.B
HS	HS-PS1-2. Construct and revise an explanation for the outcome of a simple chemical reaction based on the outermost electron states of atoms, trends in the periodic table, and knowledge of the patterns of chemical properties.	PS1.A PS1.B
	HS-PS2-3. Apply scientific and engineering ideas to design, evaluate, and refine a device that minimizes the force on a macroscopic object during a collision. *	PS2.A ETS1.A ETS1.C
	HS-PS2-5. Plan and conduct an investigation to provide evidence that an electrical current can produce a magnetic field and that a changing magnetic field can produce an electrical current.	PS2.B PS3.A
	HS-PS3-2. Develop and use models to illustrate that energy at the macroscopic scale can be accounted for as a combination of energy associated with the motions of particles (objects) and energy associated with the relative positions of particles (objects).	PS3.A

**PEs marked with an asterisk incorporate Engineering constructs.*

Table D-9. Disciplinary Core Idea Coverage for Physical Sciences Across Grades 5, 8, and HS

	Physical Sciences Disciplinary Core Idea (DCI) Coverage Across Grades												
	PS1.A	PS1.B	PS1.C	PS2.A	PS2.B	PS2.C	PS3.A	PS3.B	PS3.C	PS3.D	PS4.A	PS4.B	PS4.C
Grade 5	X	X		X	X			X		X			
Grade 8	X	X		X				X			X	X	
Grade HS	X	X		X	X		X						

Ultimately, the selected PEs within and across grades represent a content progression supporting essential learning and understandings in the Physical Sciences. The constructs of structure and properties of matter, chemical reactions, forces and motion, types of interactions, and conservation and transfer of energy are all well-represented across the grade bands. Additionally, basic understanding of waves and their behavior is included in grade 8.

Table D-10. Selected Performance Expectations for Life Science Across Grades 5, 8, and HS

Grade	Performance Expectation (PE)	DCI
5	5-PS3-1. Use models to describe that energy in animals' food (used for body repair, growth, and motion, and to maintain body warmth) was once energy from the sun.[1]	PS3.D LS1.C
	4-LS1-1. Construct an argument that plants and animals have internal and external structures that function to support survival, growth, behavior, and reproduction.	LS1.A
	3-LS3-1. Analyze and interpret data to provide evidence that plants and animals have traits inherited from parents and that variation of these traits exists in a group of similar organisms.	LS3.A LS3.B
	3-LS4-1. Analyze and interpret data from fossils to provide evidence of the organisms and environments in which they lived long ago.	LS4.A
8	MS-LS1-3. Use argument supported by evidence for how the body is a system of interacting sub-systems composed of groups of cells.	LS1.A
	MS-LS1-5. Construct a scientific explanation based on evidence for how environmental and genetic factors influence the growth of organisms.	LS1.B
	MS-LS2-1. Analyze and interpret data to provide evidence for the effects of resource availability on organisms and populations of organisms in an ecosystem.	LS2.A
	MS-LS2-3. Develop a model to describe the cycling of matter and flow of energy among living and non-living parts of an ecosystem.	LS2.B
HS	HS-LS2-2. Use mathematical representations to support and revise explanations based on evidence about factors affecting biodiversity and populations in ecosystems of different scales.	LS2.A
	HS-LS3-1. Ask questions to clarify relationships about the role of DNA and chromosomes in coding the instructions for characteristic traits passed from parents to offspring.	LS3.A LS1.A
	HS-LS4-1. Communicate scientific information that common ancestry and biological evolution are supported by multiple lines of empirical evidence.	LS4.A
	HS-LS4-3. Apply concepts of statistics and probability to support explanations that organisms with an advantageous heritable trait tend to increase in proportion to organisms lacking this trait.	LS4.B LS4.C

[1]While this PE is coded to Physical Sciences, it crosses Physical Sciences and Life Sciences. It is being counted in Life Sciences in fulfilling the blueprint distribution, as the construct would most frequently be taught in an ecology unit.

Table D-11. Disciplinary Core Idea Coverage for Life Sciences Across Grades 5, 8, and HS

Life Sciences Disciplinary Core Idea (DCI) Coverage Across Grades														
	LS1.A	LS1.B	LS1.C	LS1.D	LS2.A	LS2.B	LS2.C	LS2.D	LS3.A	LS3.B	LS4.A	LS4.B	LS4.C	LS4.D
Grade 5	X		X						X	X	X			
Grade 8	X	X			X	X								
Grade HS	X				X				X		X	X	X	

The selected PEs within and across grades represent a content progression supporting essential learning and understandings in the Life Sciences. While some DCIs get more emphasis in some grade bands than in others, the specific PEs selected for the assessments help to establish big ideas that bridge students across DCI areas as they progress through the grades. For example, in the grade 8 test, although no PEs aligned to DCI LS3 (Heredity) are assessed, the PE MS-LS1-5 includes genetic factors and thus still gives students exposure to that topic between the elementary and high school grades where specific PEs aligned to DCI LS3 are included.

Table D-12. Selected Performance Expectations for Earth and Space Science Across Grades 5, 8, and HS

Grade	Performance Expectation (PE)	DCI
5	5-ESS1-2. Represent data in graphical displays to reveal patterns of daily changes in the length and direction of shadows, day and night, and the seasonal appearance of some stars in the night sky.	ESS1.B
	3-ESS2-1. Represent data in tables and graphical displays to describe typical weather conditions expected during a particular season.	ESS2.D
	5-ESS2-1. Develop a model using an example to describe ways in which the geosphere, biosphere, hydrosphere, and/or atmosphere interact.	ESS2.A
	5-ESS3-1. Obtain and combine information about ways individual communities use science ideas to protect the Earth's resources and environment.	ESS3.C
8	MS-ESS1-1. Develop and use a model of the Earth-sun-moon system to describe the cyclic patterns of lunar phases, eclipses of the sun and moon, and seasons.	ESS1.A ESS1.B
	MS-ESS2-2. Construct an explanation based on evidence for how geoscience processes have changed Earth's surface at varying time and spatial scales.	ESS2.A ESS2.C
	MS-ESS2-4. Develop a model to describe the cycling of water through Earth's systems driven by energy from the sun and the force of gravity.	ESS2.C
	MS-ESS3-3. Apply scientific principles to design a method for monitoring and minimizing a human impact on the environment.*	ESS3.C ETS1.A
HS	HS-ESS1-6. Apply scientific reasoning and evidence from ancient Earth materials, meteorites, and other planetary surfaces to construct an account of Earth's formation and early history.	ESS1.C
	HS-ESS2-4. Use a model to describe how variations in the flow of energy into and out of Earth's systems result in changes in climate.	ESS1.B ESS2.A ESS2.D
	HS-ESS2-5. Plan and conduct an investigation of the properties of water and its effects on Earth materials and surface processes.	ESS2.C
	HS-ESS3-4. Evaluate or refine a technological solution that reduces impacts of human activities on natural systems.*	ESS3.C ETS1.B

Table D-13. Disciplinary Core Idea Coverage for Earth and Space Sciences Across Grades 5, 8, and HS

Earth and Spaces Sciences Disciplinary Core Idea (DCI) Coverage Across Grades												
	ESS1.A	ESS1.B	ESS1.C	ESS2.A	ESS2.B	ESS2.C	ESS2.D	ESS2.E	ESS3.A	ESS3.B	ESS3.C	ESS3.D
Grade 5		X		X			X				X	
Grade 8	X	X		X		X					X	
Grade HS		X	X	X		X	X				X	

The selected PEs within and across grades represent a content progression supporting essential learning and understandings in the Earth and Space Sciences. Fundamentals about Earth and the solar system lead to studying the history of planet Earth; understanding of Earth materials and systems anchors constructs in weather and climate as well as the role of water across the grades; and an emphasis on human impacts on Earth systems in all grade bands can allow for incorporation of natural resource and global climate considerations (ESS3.A, ESS3.D). DCI ESS3.B (Natural Hazards) was purposely excluded due to sensitivity concerns for this population of students.

1	NGSS Performance Expectation 5-PS2-1	
	5-PS2-1 Support an argument that the gravitational force exerted by Earth on objects is directed down. [Clarification Statement: "Down" is a local description of the direction that points toward the center of the spherical Earth.] [Assessment Boundary: Assessment does not include mathematical representation of gravitational force.]	
	Science and Engineering Practices (SEP)	**Engaging in Argument from Evidence** ● Support an argument with evidence, data, or a model.
	Disciplinary Core Ideas (DCI)	**PS2.B: Types of Interactions** ● The gravitational force of Earth acting on an object near Earth's surface pulls that object toward the planet's center.
	Crosscutting Concepts (CCC)	**Cause and Effect** ● Cause and effect relationships are routinely identified and used to explain change.

	Extended Performance Expectation 5-PS2-1			
		Level 1	Level 2	Level 3
		Less Complex		*More Complex*
2		**5-PS2-1.1** Use observations to identify patterns in the motion of objects when they are released on Earth.	**5-PS2-1.2** Select or complete a model that shows the direction objects move when they are released on Earth (downward).	**5-PS2-1.3** Describe observations, data, or a model that supports the claim that Earth's gravity pulls objects down (toward Earth's center).
3	**Science and Engineering Practices (SEP)**	**Engaging in Argument from Evidence** ● Support an argument with evidence, data, or a model. **Supporting:** **Planning and Carrying Out Investigations** **Developing and Using Models**		
	Disciplinary Core Ideas (DCI)	**PS2.B: Types of Interactions** ● The gravitational force of Earth acting on an object near Earth's surface pulls that object toward the planet's center.		
	Crosscutting Concepts (CCC)	**Cause and Effect** ● Cause and effect relationships are routinely identified and used to explain change. **Supporting: Patterns**		

1. The first box displays the general education **PE** from which the EPEs are extended. The general education PE code and text, any clarification statements and assessment boundaries, and the three dimensions of the general education PE are provided.
2. The second box displays the **EPEs at Levels 1, 2, and 3**. The coding and text of each level is presented in progression at the top of the box.
3. Information about the alignment of the **dimensions** is provided below the EPE progression. For each dimension, the ***Target*** alignment is identified. These target dimension alignments reflect the specific SEP, DCI, and CCC incorporated in the Level 3 EPE and the general education PE, and further detail the type of knowledge, skills, and abilities that students will need to apply in an integrated way to demonstrate proficiency with the Level 3 EPE. Additionally, ***Supporting*** alignments may be identified for one or more of the dimensions. These supporting alignments list additional SEPs or CCCs that may be incorporated at the Level 1 and/or Level 2 access points to scaffold student learning towards Level 3 proficiency.

APPENDIX—E
VARIABLE FEATURES AND SUPPORTS

Target Volume of Information (VI)

The amount of information to be contained in the stimulus/scenario for an item is detailed in the specifications as “Target Volume of Information.” The chart below provides the key to the 1-4 coding presented in the specifications.

	1	2	3	4
Volume of Information (VI)	**No Scenario Presented:** • 1 simple sentence stating stimulus, “This is a ….” or “The picture shows ….” (when applicable) • Little to no additional info or instruction beyond standard item template language • Minimal response options (no complete sentences or equations) • No passage *Which picture shows an animal and its baby?* (no stimulus, 3 pictures)	**Limited Scenario Presented:** • 1 sentence describing stimulus or scenario • Minimal information provided in 1 simple format (graph, diagram, organizer, formula) • May have no scenario, but response options may be complete sentences or equations *Jose makes a model to show part of the water cycle.* *[model graphic]* *What is one path of the water as it moves through the water cycle in the model?* (1 information sentence and model in stimulus, 3 sentences)	**Moderate Scenario Presented:** • 2 sentences describing stimulus or scenario • Moderate information provided in 1 format (graph, data table, diagram, organizer, formula) *John is studying how Earth is warmed by the Sun. He made a model to show how the Sun’s energy warms the Earth.* *[model graphic]* *Based on the model, what is one pathway for energy flow between Earth systems?* (2 information sentences and model in stimulus, 3 word card sequences)	**Complex Scenario Presented:** • 3 or more sentences describing stimulus or scenario • Extensive information provided in 1 or more formats (graph, data table, diagram, chart, organizer, formula) *Melissa uses a model that shows the energy transfers that occur in a roller coaster. The information that Melissa records is shown in the data table. Melissa claims that the kinetic energy of the roller coaster changes as it moves along the track.* *[data table]* *Which data supports Melissa’s claim?* (multiple information sentences and data table in stimulus, 3 sentences)

Note that Level 1 items may or may not contain a stimulus. Level 2 and Level 3 items will almost always contain a stimulus. The volume of information being targeted in the stimulus will be dependent on the specific language and expectation of the EPE access point, as reflected in the specifications for the EPE.

In all cases, attend to universal design principles, using short, clear sentences and simple pictures/tables/graphics.

Contexts

Almost all items should present a specific context for the questions. Contexts may vary from being very simple, familiar, and “local” to being more complex, less familiar in content and setting, and sometimes even abstract. The chart below provides the key to the 1-4 coding presented in the specifications.

	1	2	3	4
Context (C)	**Familiar Context & Immediate Setting** (home and school) Contexts: schedule, lunch, recess, counting objects, weather, basic body parts, gravity on everyday objects, measuring height of everyday objects, family *Alternatively, no context provided Settings: home, classroom, media center, kitchen	**Familiar Context & Immediate Setting** (community) Contexts: volunteering, familiar animals/facts, more complex procedures (e.g., measuring weight before and after mixing or heating) Settings: town library/museum, grocery store, local parks and streams, well known environments (forest, farm, desert)	**Unfamiliar Context & Extended Setting** (global community) Contexts: animals, life cycles, respiratory system, internal functions of organs, Settings: Olympics, national parks and wildlife refuges, large rivers, oceans, clouds, volcanoes	**Unfamiliar & Abstract Context** (require student to apply knowledge) Contexts: carbon cycle, gene inheritance, glucose production in photosynthesis, gravity in space, changes in Earth’s position relative to Sun, periodic table, particles, molecules, model systems (stream tables)

The item specifications also provide sample context to illustrate the types of contexts and scenarios that could work well with each particular access point. The contexts of actual items will vary beyond those shown in the sample items.

Target Vocabulary

Vocabulary used in alternate assessment items must be chosen carefully. The item specifications for each EPE contain suggestion for allowable and not-allowable vocabulary specifically related to the access points. As vocabulary may vary item by item based on the specific context and focus of the item, the specifications also contain a more generalized target for the level of vocabulary to be used in items at each level of the EPE. The chart below provided the key to the 1-4 coding present in the specifications.

	1	2	3	4
Vocabulary (V)	**Familiar Vocabulary Presented:** • Everyday words and single digit numbers (e.g., round shape, “which is a boy?”, “which is wet?” presented in item) • Examples of basic content words include food, rain, Sun, hot, push, pull	**Somewhat Familiar Vocabulary Presented:** • Everyday words and basic content words used • Basic content words are words with science meaning that are used in conversation • Examples of basic content words include units of measure, data tables, graphs, decimals, light, time, gravity, electricity, energy, pattern	**Familiar & Unfamiliar Vocabulary Presented:** • Mix of everyday words and unfamiliar words • More specialized content words used • Examples of more specialized content words include laboratory tools, predict, effect, resource, density, precipitation, evaporation, population	**Abstract & Unfamiliar Vocabulary Presented:** • Abstract and/or complex content words introduced • Examples of abstract and/or complex content words include evidence, claim, carbon cycle, chlorophyll, carbon dioxide, atom, respiration

**Note that as grade level increases, familiarity with words may also increase. Therefore, a word that is unfamiliar at grade 5 may be familiar at grade 8 or HS.*

In all cases, attend to universal design principles, using the most familiar and basic vocabulary possible without sacrificing content accuracy; using consistent vocabulary through an item’s stimulus text, graphic, stem, and options; and clearly defining unfamiliar words when it is necessary to include them.

APPENDIX—F
TEST DESIGN BLUEPRINTS

NGSS-Alt Blueprints by Grade

Grade 5				
PE	Text	DCI	SEP	CCC
5-PS1-2	Measure and graph quantities to provide evidence that regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is conserved.	1.A, 1.B	Using Mathematics and Computational Thinking	Scale, Prop, Quantity
3-PS2-2	Make observations and/or measurements of an object's motion to provide evidence that a pattern can be used to predict future motion.	2.A	Planning and Carrying Out Investigations	Patterns
5-PS2-1	Support an argument that the gravitational force exerted by Earth on objects is directed down.	2.B	Engaging in Argument From Evidence	Cause and Effect
4-PS3-4	Apply scientific ideas to design, test, and refine a device that converts energy from one form to another.*	3.B, 3.D ETS1.A	Constructing Explanations, Designing Solutions	Energy and Matter
5-PS3-1	Use models to describe that energy in animals' food (used for body repair, growth, and motion, and to maintain body warmth) was once energy from the sun.[1]	3.D, (LS1.C)	Developing and Using Models	Energy and Matter
4-LS1-1	Construct an argument that plants and animals have internal and external structures that function to support survival, growth, behavior, and reproduction.	1.A	Engaging in Argument from Evidence	Systems and System Models
3-LS3-1	Analyze and interpret data to provide evidence that plants and animals have traits inherited from parents and that variation of these traits exists in a group of similar organisms.	3.A, 3.B	Analyzing and Interpreting Data	Patterns
3-LS4-1	Analyze and interpret data from fossils to provide evidence of the organisms and environments in which they lived long ago.	4.A	Analyzing and Interpreting Data	Scale, Prop, Quantity
5-ESS1-2	Represent data in graphical displays to reveal patterns of daily changes in the length and direction of shadows, day and night, and the seasonal appearance of some stars in the night sky.	1.B	Analyzing and Interpreting Data	Patterns
3-ESS2-1	Represent data in tables and graphical displays to describe typical weather conditions expected during a particular season.	2.D	Analyzing and Interpreting Data	Patterns
5-ESS2-1	Develop a model using an example to describe ways in which the geosphere, biosphere, hydrosphere, and/or atmosphere interact.	2.A	Developing and Using Models	Systems and System Models
5-ESS3-1	Obtain and combine information about ways individual communities use science ideas to protect the Earth's resources and environment.	3.C	Obtaining, Evaluating, and Communicating Information	Systems and System Models

**PEs marked with an asterisk incorporate Engineering constructs.*

[1]This PE crosses Physical Sciences and Life Sciences. It will be classified in Life Sciences in fulfilling the blueprint distribution.

Grade 8				
PE	**Text**	**DCI**	**SEP**	**CCC**
MS-PS1-2	Analyze and interpret data on the properties of substances before and after the substances interact to determine if a chemical reaction has occurred.	1.A, 1.B	Analyzing and Interpreting Data	Patterns
MS-PS2-2	Plan an investigation to provide evidence that the change in an object's motion depends on the sum of the forces on the object and the mass of the object.	2.A	Planning and Carrying Out Investigations	Stability and Change
MS-PS3-5	Construct, use, and present arguments to support the claim that when the kinetic energy of an object changes, energy is transferred to or from the object.	3.B	Engaging in Argument From Evidence	Energy and Matter
MS-PS4-2	Develop and use a model to describe that waves are reflected, absorbed, or transmitted through various materials.	4.A, 4.B	Developing and Using Models	Structure and Function
MS-LS1-3	Use argument supported by evidence for how the body is a system of interacting sub-systems composed of groups of cells.	1.A	Engaging in Argument from Evidence	Systems and System Models
MS-LS1-5	Construct a scientific explanation based on evidence for how environmental and genetic factors influence the growth of organisms.	1.B	Constructing Explanations, Designing Solutions	Cause and Effect
MS-LS2-1	Analyze and interpret data to provide evidence for the effects of resource availability on organisms and populations of organisms in an ecosystem.	2.A	Analyzing and Interpreting Data	Cause and Effect
MS-LS2-3	Develop a model to describe the cycling of matter and flow of energy among living and non-living parts of an ecosystem.	2.B	Developing and Using Models	Energy and Matter
MS-ESS1-1	Develop and use a model of the Earth-sun-moon system to describe the cyclic patterns of lunar phases, eclipses of the sun and moon, and seasons.	1.A, 1.B	Developing and Using Models	Patterns
MS-ESS2-2	Construct an explanation based on evidence for how geoscience processes have changed Earth's surface at varying time and spatial scales.	2.A, 2.C	Constructing Explanations, Designing Solutions	Scale, Prop, Quantity
MS-ESS2-4	Develop a model to describe the cycling of water through Earth's systems driven by energy from the sun and the force of gravity.	2.C	Developing and Using Models	Energy and Matter
MS-ESS3-3	Apply scientific principles to design a method for monitoring and minimizing a human impact on the environment.*	3.C (ETS1.A)	Constructing Explanations, Designing Solutions	Cause and Effect

**PEs marked with an asterisk incorporate Engineering constructs.*

HIGH SCHOOL				
PE	Text	DCI	SEP	CCC
HS-PS1-2	Construct and revise an explanation for the outcome of a simple chemical reaction based on the outermost electron states of atoms, trends in the periodic table, and knowledge of the patterns of chemical properties.	1.A, 1.B	Constructing Explanations, Designing Solutions	Patterns
HS-PS2-3	Apply scientific and engineering ideas to design, evaluate, and refine a device that minimizes the force on a macroscopic object during a collision.*	2.A ETS1.A, ETS1.C	Constructing Explanations, Designing Solutions	Cause and Effect
HS-PS2-5	Plan and conduct an investigation to provide evidence that an electrical current can produce a magnetic field and that a changing magnetic field can produce an electrical current.	2.B, (3.A)	Planning and Carrying Out Investigations	Cause and Effect
HS-PS3-2	Develop and use models to illustrate that energy at the macroscopic scale can be accounted for as a combination of energy associated with the motions of particles (objects) and energy associated with the relative positions of particles (objects).	3.A	Developing and Using Models	Energy and Matter
HS-LS2-2	Use mathematical representations to support and revise explanations based on evidence about factors affecting biodiversity and populations in ecosystems of different scales.	2.A	Using Mathematics and Computational Thinking	Scale, Prop, Quantity
HS-LS3-1	Ask questions to clarify relationships about the role of DNA and chromosomes in coding the instructions for characteristic traits passed from parents to offspring.	3.A, (1.A)	Asking Questions	Cause and Effect
HS-LS4-1	Communicate scientific information that common ancestry and biological evolution are supported by multiple lines of empirical evidence.	4.A	Obtaining, Evaluating, and Communicating Information	Patterns
HS-LS4-3	Apply concepts of statistics and probability to support explanations that organisms with an advantageous heritable trait tend to increase in proportion to organisms lacking this trait.	4.B, 4.C	Analyzing and Interpreting Data	Patterns
HS-ESS1-6	Apply scientific reasoning and evidence from ancient Earth materials, meteorites, and other planetary surfaces to construct an account of Earth's formation and early history.	1.C	Constructing Explanations, Designing Solutions	Stability and Change
HS-ESS2-4	Use a model to describe how variations in the flow of energy into and out of Earth's systems result in changes in climate.	1.B, 2.A, 2.D	Developing and Using Models	Cause and Effect
HS-ESS2-5	Plan and conduct an investigation of the properties of water and its effects on Earth materials and surface processes.	2.C	Planning and Carrying Out Investigations	Structure and Function
HS-ESS3-4	Evaluate or refine a technological solution that reduces impacts of human activities on natural systems.*	3.C ETS1.B	Constructing Explanations, Designing Solutions	Stability and Change

**PEs marked with an asterisk incorporate Engineering constructs.*

APPENDIX G
REPORTING SERVICES DELIVERABLES DECISION RULES

MSAA Assessments Reporting Services Deliverables Decision Rules

2022-2023

02/09/2023

Tara LaPierre

cognia™

Table of Contents

Overview

This document describes the Reporting Services administration analysis and reporting requirements for the Multi-State Alternate Assessment (MSAA) administered during the **2022-2023** academic school year. For each Reporting Services responsibility, information needed to produce accurate and timely deliverables is included throughout this document.

Partners

MSAA is a consortium of Partners. Each Partner may select various analysis and reporting deliverable options. The active Partner for the current school year is included in the table below.

Partner	Partner Abbreviation
American Samoa	AS
Arizona	AZ
Bureau of Indian Education (BIE)	BI
District of Columbia	DC
Department of Defense Education Activity (DoDEA)	DD
Guam	GU
Maine	ME
Commonwealth of Northern Mariana Islands (CNMI)	MP
Montana	MT
South Dakota	SD
Tennessee	TN
Vermont	VT
US Virgin Islands	VI

Assessment and Administrations

The MSAA contract consists of ELA and mathematics assessments administered during the school year to grades 03-08, and 11. As a Partner option, Science may also be administered during the school year to grades 05, 08, and 11 students. Breakthrough's system will be used for registration and administration of the assessments. Student test data will be collected online only; there will be no scannable documents.

Assessment Content Area	Assessment Grade*	Brief Description	Start Date	End Date	Partner
English Language Arts	03	Stage-Adaptive test that includes operational and embedded field test items consisting of Single Select Choice Items and Writing Prompt item types	3/13/2023	04/28/2023	All
	04				
	05				
	06				
	07				
	08				
	11				
Mathematics	03	Stage-Adaptive test that includes operational and embedded field test items consisting of Single Select Choice Items	3/13/2023	04/28/2023	All
	04				
	05				
	06				
	07				
	08				
	11				
Science	05	Operational Field test consisting of Single Select Choice Items	3/13/2023	04/28/2023	AS, AZ, BI, MP, GU, ME, VI, VT
	08				
	11				

(*) VT will administer the grade 11 ELA and mathematics tests to grade 09 students only.

Reporting Services Deliverables List

Reporting Services will produce various data file and static report deliverables included in the table below. This document details the data preparation, processing, and formatting rules.

Post-Test Administration Deliverable		Partners
Student Demographics Datafile (for Test Clean-Up)		All
Organization Datafile (for Test Clean-Up)		All
Test Materials Download Count		All
Writing Score Off-Topic		All
Billable Records Datafile (True-Up File)		All
Scaled Score Lookup Datafiles		AZ
Student Results School, District, and State Datafiles	sFTP Preliminary (State Only)	All
	BT Online Final (School, District, State)	All
	Focal Point Online Final (State)	VT
	sFTP Final (State Only)	All
Duplicate/Void State Student Test Datafiles	sFTP Final	All

Post-Test Administration Deliverable		Partners
Student Report	BT Online	All
	eMetric Online	BI
	Print	TN, BI
School and District Roster Report	BT Online	All
School, District, and State Summary Report	BT Online	All
eMetric Data Interaction (DI)	eMetric Online	BI
Parental Rescore Request		SD

Change Log

Administration	Description
• 2020-2021	• Datafile deliverables will be in EXCEL format instead of CSV
• 2020-2021	• ELA/Math Student Roster will be created at the district level in addition to school
• 2020-2021	• Any Partner choosing the Student Report Print Option will receive two copies of the report
• 2020-2021	• When both/all tests for a student are not launched/started but are closed (due to TA/TC misadministration) will be reported as ESM. These tests were previously reported as DNT.
• 2020-2021	• Science will be administered to the Partners who select the science option • Science Participation file will be created after test clean-up
• 2020-2021	• Administration window extended from 04/30/2021 to 05/14/2021 • Note: SD admin ends 05/07/2021
• 2020-2021	• Student Demographic test clean-process modified by combining bull-pen and demographic process • Partners can provide information for Cognia to add, remove, merge student data to be included in analysis and reporting • Partners can provide information for Cognia to update demographics, test status, and reporting status (participation status) • The process is outlined in the requirements document MSAA 2122 Student Demographic Instructions.pdf • Final reporting status values will be calculated as part of the demographic clean-up process and detailed in the requirements document MSAA 2122 Student Demographic Instructions.pdf
• 2020-2021	• DC does not plan to administer MSAA in 2021
• 2020-2021	• Do not print the Scaled Score Low/High sentence on the student report for students with a reporting status of ESR
• 2020-2021	• WRP Reporting Status will stay in the student results file. However, the rules will be to submit a value if different from ELA reporting status and blank if the same. Change the valid values to remove those that would not apply (remove TES, ESR, ESM, INC, ELL, EXE, DNT, WDR, and NLE).

Administration	Description
	• A few supports/accommodation fields were removed from the student results layout since they no longer exist
• 2021-2022	• Science will be operational. Two major rounds of reports and datafiles: Pre-Standard Setting and Post-Standard Setting. Pre-Standard Setting will include ELA/Math Results and Science Participation as outlined in this document. Post-Standard Setting will include ELA/Math/Science results for Partners who participated in the Science assessment. • DoDEA joined MSAA • BI will have eMetric Data Interaction Reporting • Science Student Report design • School and District Student Roster re-designed such that one subject is reported on a single page. • Print Ready Student Report PDFs for Partners who opted in • Student Results Layout for ELA/Math/Sci new – modeled off of ELA/Math layout • Added calculation rule for “Ethnic” using the individual Race/Ethnicity variables to create one Ethnic variable
• 2022-2023	• VT joined MSAA ELA, Math, and Science: Testing at grade 09 instead of grade 11 for ELA and Math. The grade 09 ELA and mathematics tests will be analyzed as test grade 11 for psychometric analyses. • Additional partners are participating in Science: AS, AZ, BI, MP, GU, ME, VI, VT • BT organization ID management /assignment and creation changed. BT Org ID should not change across years. (No impact on Reporting.) • Standards Setting in Science • No more Print Ready PDFs • Two new accommodations (SAR_Braille_Before and SAR_Braille_After) added to student results layout • Added Grade 09 as a valid value to the student results layout • Reporting will only use the final BT extract and ignore the initial BT extract for analysis reporting • Static Reporting: All reports - “Grade 11” will be replaced with “High School”; Remove science provisional score footnote; Student Report – parent letter edits and What to work on next text updates • Focal Point will receive VT state student results file for reporting and summary aggregation files for QC. File names for eMetric changed.

Pre-Test Administration Data Preparation

Organizational Data

Partners Cognia Operational Services department district and school data following a standardized layout. Cognia will load the data into an internal database referred to as ICORE. The requirements for district and school organizational handoff, load into ICORE, and data maintenance is out of scope for this document. However, the data will be used to support reporting assessment results. Internal use only school and district organizations are added to ICORE to support quality assurance. The fields and value descriptions used for MSAA reporting are detailed below.

MSAA Reporting Organizational Data Descriptions

Field	Field Description
ReportCode1	• Partner Abbreviation • Partner code DEMO are for internal use only
BT Org ID	• Unique code assigned by the Breakthrough Portal to identify the Partners, Districts, and Schools
District Code	• Unique code (within Partner) to identify districts • District Code values of DEMOA and DEMOB are for internal use only • Length and Pattern of Values Varies
District Name	• District name used for reporting • ASCII Text field • Maximum allowable length 30
School Code	• Unique code (within Partner) when combined with District Code identifies a unique School • Schools associated with District Code values DEMOA and DEMOB are for internal use only • Length Varies and Pattern of Values
School Name	• School Name used for reporting • ASCII Text field • Maximum allowable length 30

MSAA ICORE Data Store

ICORE contract code is used to identify the set of organizational data used to support analysis and reporting.

Administration	ICORE Contract Code	Partners
Spring 2023	• 603200, 603250, 603252	• All

Test Meta Data

The information in this section describes the test meta data needed to support data student test data validation as well as analysis and reporting activities. Test meta data includes information about tests, forms, and items being administered. Test meta data impacting analysis and reporting include Test Form ID, Test Form Session & Position, Item Number, Item Type, Item Points, Item Subject, Count Towards Student Score, Item Role on Test Form, Equating Eligible Status.

Source

NTS is the primary test meta-data source support MSAA analysis and reporting. Test meta data will be extracted from NTS after Content Development and Publications Cognia department (CDP) completes test clean up.

Session Forms

MSAA is designed to be stage adaptive. The student's score on the first session determines what form will be administered in the second session. Therefore, Forms will be constructed at the session level. Each eligible student is expected to take one form for session 1 and one form for session 2 for an assessment content area (also referred to as test). All forms will be available in English only. Note: Science is not Stage-Adaptive in 2022-2023.

Test Session & Position

Within the NTS data, for Stage-Adaptive tests, each form consists of one session where each session consists of a collection of items. The NTS form name includes the session. The position field indicates the order items are presented to students. Position should be unique on a test form.

Special Processing of Form Meta Data

Session Form data will be used to create Test Form data by combining all possible combinations of Session 1 Form and Session 2 Forms.

Item Number

Item number (NTS AssetID) is used to support various psychometric analyses as well linking student test data to NTS data.

Item Types

Each item is characterized by its type. The item type identifies student response and score data formats. The table below lists the item types administered by MSAA. MSAA tests consist of single select choice items and a writing prompt (ELA only). Writing prompts are scored on three trait dimensions: Organization, Idea Development, and Conventions.

Item Type Label	NTS Identification	Reporting Abbreviation	CDP Abbreviation
Single-Select Choice	Interaction Type: choiceInteraction, and Correct Response: Exactly one option is the correct response	MC	SR
Writing Prompt: Scored on Three Dimensions/Traits	PointValue = 9 (Note: Each Dimension scored on 3 points)	WP	WP

Item Role on Test Form

Each item on a form is characterized as operational or field test. An item's role on a test form impacts various analyses including calculating student test scores.

Role	Abbreviation	Rule
Operational	OP	• Included in calculating student test scores • CountsTowardStudentScore = Yes
Field Test	FT	• Excluded from calculating student test scores • CountsTowardStudentScore = No

Stage-Adaptive Requirements

Reporting provides Psychometrics session 1 scaling items item lists. Psychometrics provides the routing lookups to Cognia CDP department to be incorporated in test production. Psychometrics determines the raw scores for each session 1 form required for session 2 form assignment. Since Science is not stage-adaptive in 2022-2023, routing item lists will not be produced.

ELA Reading and Writing Items

Every ELA item is assigned a Subject value of Reading or Writing in NTS. The Subject code is used for calculating Reading Percent of Points Earned and Writing Percent of Points Earned.

Test Administration Validation

Reporting participates in validating Breakthrough MSAA Testing System prior to the system going live for an administration.

Student Registration Data

Student registration occurs with each Partner utilizing the Breakthrough MSAA Systems Portal. Registration requirements are outside the scope of this document. Each student will be associated to a Partner, district within the Partner, and school within the Partner in the portal.

Post-Test Administration Data Clean-up

Report Services receives data from various sources, validates the data, and applies processing rules to prepare data for psychometrics, analysis, and report generation. This section provides a general overview of the various sources and a detailed description of student item responses and scores as well as test status. In-depth detail on the data processing rules and data sources are out of scope of this document.

Student Data Sources

Student Online Test Data – BT Systems Portal	
Description	• Breakthrough will provide Cognia data related to student online testing following and agreed upon schedule. • The data includes o student demographics at the time of testing, o student accommodation, o LCI data, o student response check data, o student test data including not tested reasons, student test session data, test date time stamp, student item responses item evidence, and scores, o test meta data o test proctor data, o organization data o student's at the '9999' organization are excluded
General Rules	• Cognia Reporting will import and validate the files • Cognia Reporting will provide item evidence counts to Cognia Client Services for conformation that all evidence files have been received for scoring
File Layout	• BT provides Cognia standard CSV files following an agreed upon format

Demographic File – Partner Updated	
Description	• Partners provide an updated student demographic data file • Cognia will incorporate updates as part of post-test administration student test cleanup
General Rules	• Refer to MSAA 2223 Student Demographic Instructions.pdf
File Layout	• Refer to MSAA 2223 Student Demographic Instructions.pdf

Student Human Item Scores	
Description	• Cognia Scoring Services will provide Reporting Services student level item scores and non-scorable scores
General Rules	• Refer to section “Student Item Response: Human Score Type “
File Layout	• Scoring Specifications

Student Item Data

The purpose of this section is to describe in detail the data associated with items on student tests necessary for analysis and reporting and student data clean-up activities.

Student Item Response: Format

Student item responses are captured and formatted and stored as described below. Item type is used to categorize the response formats.

Item Type	Student Response Description	Sample Value
Single-Select Choice	• Single alpha character	• A
Writing Prompt	• N/A	N/A

Student Item Response: Scoring Method

Each student response to an item is assigned a score value. An item score is assigned either by machine scored or human scored. Student responses collected online is either machine scored by the testing platform or human scored.

Item Type	CBT	PBT	Scoring
Single-Select Choice	Testing Platform	N/A	Exact Match: 1 = student response match correct response; 0 otherwise
Writing Prompt	Human	N/A	Refer to sections Student Item Response: Human Score Type, Writing Prompt: Valid Dimension Score Combinations and Writing Prompt: Score Adjustment sections below

Item Excluded: Identify Student Modified Test Form

Rarely an administration issue may lead to excluding an item from a student test form during test clean-up. To exclude the item from scoring a particular student’s test, the item response is set to X and score set to blank.   Student test scores will be based on all core items administered the student where the response is not X.

Writing Prompt: Raw Trait Dimension Scores

Student responses requiring a human score will have a final score of record, scorer 1 score, scorer 2, and scorer 3 score as defined by scoring procedures. The final score of record value is used to calculate official student test scores and used to determine if a student attempted an item. Refer to the Writing Prompt: Score Adjustment section for more information on the wring prompt score. Scoring rubrics and procedures are out of scope for this document. Each student response requiring a human score will be assigned a final score of record score value for each rubric dimension as outlined in the table below.

Human Score	Interpretation	Raw iScore Value	Valid*	Item Attempt**
Numeric	Valid numeric score (an integer greater than or equal to 0 and less than or equal maximum allowed item score as defined in the rubric)	0,1,2,3	OP, FT	Yes
Blank	No deliberate marks in the answer space; No evidence submitted	B	OP, FT	No
Unreadable	Faint handwriting or otherwise obstructed student response	U	FT	Yes
Non-English	Response is written in a language other than English, or is a mix of English and another language but lacks sufficient English to provide a score	F	OP, FT	Yes
Off Topic	A response that is not related to the task/prompt administered or is not a valid attempt at answering any task/prompt on the test	5	OP, FT	Yes
Repeats the Prompt	The response copies the prompt or portions of it and offers no attempt to respond to the task/prompt	P	OP, FT	Yes
No Score	Any other response that cannot receive a numeric score	N	OP, FT	Yes
Insufficient Amount to Score	The response contains an insufficient amount of writing to score	A	N/A	Yes
Refusal	The response clearly indicates a refusal on the part of the student to address the prompt or participate in the test	R	N/A	Yes
Illegible	Tiny or poor handwriting, spelling that cannot be deciphered, or other conditions that render the student work indecipherable	I	N/A	Yes
Wrong Location	Item response inconsistent with student form	W	N/A	Yes
Response Not Scored	Field test item where students' response was not selected for scoring	# or blank	FT	Unknown

(*) Valid: OP = Human score value is valid for operational items
FT = Human score value is valid for field test items
N/A = Not applicable for project. If value provided, resolution needed.
Note: In 2022-2023, all Writing Prompts are OP.
(**) Item Attempt: Yes = Human score value indicates student attempted the item
No = Human score value indicates student did not attempt the item
Unknown = Not enough information to determine if the student attempted the item

Writing Prompt: Valid Trait Dimension Score Combinations

Writing prompts are scored on three trait dimensions: Organization, Idea Development, and Conventions. Each trait is assigned a score listed in the "Raw Score Value" column in "Writing Prompt: Raw Trait Dimension Scores". Off Topic is not a valid score for the Conventions trait. If one dimension score is scored a B, then all dimension scores must be a B.

Writing Prompt: Dimension Score Adjustment

The raw iScore dimension score values are translated as indicated below to support analysis and reporting requirements. During test cleanup, the raw iScore value is translated to the Student Results value except Z will be set to B to be consistent with standard processes. “B” will be translated to “Z” when producing the student results and void/duplicate files

Human Score	Raw iScore Value	Psychometric Score Value	Student Results
Rubric Score	0	0	0
Rubric Score	1	1	1
Rubric Score	2	1	1
Rubric Score	3	2	2
Blank	B	0	Z
Unreadable	U		U
Non-English	F	0	F
Off Topic	5	0	O
Repeats the Prompt	P	0	P
No Score	N	0	N
Item Excluded: Identify Student Modified Test Form during Clean Up	0-3,5, B, U, F, P, N		X

Single-Select Choice Response: Response Adjustment

Student responses to single-select choice items are translated below to support analysis and reporting.

Raw Response	Raw Value	Psychometric Score Value	Student Results
Raw Response	A, B, C, or D	0 = response does not match item key	A, B, C or D
		1 = response matches item key	+
Raw Response	blank	0	Z
Item Excluded: Identify Student Modified Test Form during Clean Up	A, B, C, D, or blank		X

Student Item Attempt

Item Type	Item Attempt Rule
Single-Select Choice	If student raw response is not blank or X, the student attempted the item
Writing Prompt	If the student’s earned score value for one or more dimensions is listed as a “Yes” in “Item Attempt” column in “Writing Prompt: Raw Trait Dimension Scores” table, the student attempted the item.

Student Test Data

Test data applies at the ELA, mathematics, and Science levels. Science test data will only exist for Partners who selected the option to administer the science test. The purpose of this section is to describe in detail the data associated with student tests necessary for analysis and reporting and student data Clean-Up activities.

Student Test Status

Each student test is assigned a test status in the Breakthrough Portal and adjusted during student data Clean-Up when necessary. This field will be updated during demographic clean-up.

Final Test Status	Condition
InProgress	• BT Portal value Paused value is changed to InProgress during test Clean-Up • Provided by field using BT Portal
Cancelled	• Provided by field using BT Portal • Canceled test status is also referred to as Closed Tests
Completed	• Provided by field using BT Portal • Completed test status value is also referred to as Submitted
[Blank]	• Final Test Status will be blank for Science if a Partner does not participate in Science • Final Test Status will be blank for students who were added during demographic clean up

Student Reporting Status (Participation Status)

Each student is assigned an ELA Reporting Status, Mathematics Reporting Status, a Writing Reporting Status, and Science Reporting Status during test cleanup. The allowed values are detailed in the table below. If a partner does not participate in Science, the Science Reporting Status will be blank. The rules for assigning the final reporting status are out of scope of this document. Refer to student demographic clean-up instructions for reporting status assignment rules.

Test Reporting Status	Code	Description
Administration Irregularity	IRR	Administration irregularity reported, but does not necessitate an invalidation
Invalidated	INV	Student-based or administration-based irregularity resulting in invalidation
Parental Refusal	PRF	Parental refusal
ELL Exempt (ELA Only)	ELL	Student meets the ELA ELL 1st Year in U.S. exemption requirements
Exempt	EXE	Student meets test exemption requirements
Withdrew	WDR	Student withdrew
No Longer Eligible	NLE	Student is no longer eligible for testing
Tested	TES	Submitted test, regardless of number of item responses
Tested-Incomplete	INC	In-Progress Test, with at least one item response
Early Stopping Rule	ESR	Closed Test – with no item response
Early Stopping Rule – Misadministration	ESM	Closed Test – with at least one item response Closed Test – both/all content area tests not launched or started
Did Not Test	DNT	No Test, or In-Progress Test with no item response

Post-Test Administration Student Data Clean-Up

Various data sources, including Test Meta Data, Organization Data, Online Student Test Data, Scores for Human scored items, and Demographic Clean-Up are used to conduct student data clean-up to produce student test data ready for analysis and reporting. The table below describes relevant detail related to the clean-up process and requirements.

Data	Guidelines
General Information	• Cognia will update student data using the updated demographic files returned by each Partner as outlined in the Demographic Clean-Up Instructions for additional details • Updates include modifying demographic, test status, preliminary reporting (participation status), item responses/scores data as well as adding and removing student tests • After the updates are incorporated, Cognia will perform additional clean up as outlined below
Organization Data	• All student test records associated with the same student ID must have the same School, District, and State • State, District, and School codes associated with student tests must exist in ICORE and Breakthrough Organization file. • New or revised Organization data will be updated in both ICORE and Breakthrough reporting platforms • Cognia will work with Partners to identify the complete set of schools and district organizations, along with the names for reporting, during the demographic file acceptance and organization Clean-Up process with each Partner
Student Test Grade	• Test grade is expected to match Student Enrolled Grade. • If a student's enrolled grade level is provided in the final demographic data does not match the student's tested grade, the test is considered off-grade and will be marked as "Void/Duplicate"
Duplicate Test	• After Off-Grade tests have been resolved, duplicate tests are tests in the same Assessed Content Area and State Student ID within a State Partner • The final test used for analysis and reporting is determined used the following hierarchy o Submitted/Completed o Closed o In Progress o If two or more tests have the same status, the test associated with the latest date will be used, determined by the datetime stamp of the test record. Additionally, the larger TestID is used if still duplicate. • The duplicate test(s) not selected for analysis and reporting will marked as "Void/Duplicate"

continued

Data	Guidelines
Student Test Status	• Final ELA, Mathematics, and Science Test Status will be audited based on MSAA 2223 Student Demographic Instructions.pdf
Student Test Reporting Status	• Final ELA, Mathematics, and Science Test Reporting Status (Participation Status) will be calculated based on MSAA 2223 Student Demographic Instructions.pdf
Student Writing Prompt Reporting Status	• Final Writing Prompt Reporting Status (Participation Status) will be calculated based on MSAA 2223 Student Demographic Instructions.pdf
Student Test Item Responses	• Item responses could be removed based on Student Test Reporting Status as detailed in the demographic clean up instructions
Ethnic	• For DIF and eMetric DI the algorithm below is applied to assign one Ethnic value as follows: • If Hispanic is indicated, then “Hispanic” • Else, if DemographicRaceTwoOrMoreRaces is indicated then “Multi” • Else if AmericanIndianOrAlaskaNative is indicated then “AIAN” • Else if Asian is indicated then “Asian” • Else if BlackorAfricanAmerican is indicated, then “BAA” • Else if NativeHawaiianOthPacificIslander is indicated, then “NHOPI”, • Else if White is indicated, then “White”

Post-Test Administration Psychometric Data

Reporting Services will provide Cognia Psychometric team test meta data and student test administration data consisting of demographics, student test status, student test form, and student item level responses and scores. Psychometrics will conduct statistical key checks, Stringer Analyses, CTT, and IRT. The specifications for such activities are out of scope for this document. Psychometrics will provide Reporting Services pre-equated test scaling information and raw score to scaled score lookup tables as described in this section to support creation of data file and report deliverables.

Psychometrics Assigned Scores	
ELA Cut Scores by Test Grade	• Proficiency Level Scale Score Ranges
ELA Scaled Score Lookup by Test Grade	• Scale form • Raw Score • Scale Score • Proficiency Level • Scale Score Low/High
Mathematics Cut Scores by Test Grade	• Proficiency Level Scale Score Ranges
Mathematics Scaled Score Lookup by Test Grade	• Scale form • Raw Score • Scale Score • Proficiency Level • Scale Score Low/High
Science Cut Scores by Test Grade	• Proficiency Level Scale Score Ranges • Available after Science Standard Setting 2022-23
Science Scaled Score Lookup by Test Grade	• Scale form • Raw Score • Scale Score • Proficiency Level • Available after Science Standard Setting 2022-23

Post-Test Administration Reporting Calculations

This section details calculations and formatting applied after test clean-up is complete.

Student Data

The data listed below details student level data used to support various analysis and reporting tasks. It does not include a complete list of student data fields available. Student data prepared for psychometrics is merged with student scores calculated by psychometrics. [Test] Refers to ELA, mathematics, and science tests. Science test fields will be blank for Partners who did not participate in science.

Field	Description
[Test] Form	• Two letter test form identification where the first letter identifies the session1 form and the second letter identifies the session 2 form • Students without a test form who need to be reported are defaulted to form AA or 01
[Test] Scale Form	• Identifies the unique set of scaling and equating items based on Test Form and “Item Excluded: Identify Student Modified Test Form during Clean Up”
[Test] Form Modified	• If during test clean up the student test was identified as “Item Excluded: Identify Student Modified Test Form during Clean Up” the field will be set to a “1” ; otherwise it will be “0”
[Test] Raw Score	• Sum of final non-flawed item scores classified as “counts toward student score” items for the student test
[Test] Scaled Score	• Using calculated Test Scale Form, Test Raw Score and Psychometric Raw Score to scale score lookup, assign a Test Scaled Score • Apply Reporting Status test score rules as appropriate for a specific deliverable
[Test] Performance Level	• Using calculated Test Scale Form, Test Raw Score and Psychometric Raw Score to scale score lookup, assign a Test Performance Level • Apply Reporting Status test score rules as appropriate for a specific deliverable
[Test] Scaled Score Low/High	• Using calculated Test Scale Form, Test Raw Score and Psychometric Raw Score to scale score lookup, assign a Test Scaled Score Low/High • Apply Reporting Status test score rules as appropriate for a specific deliverable
[Test] State Compare	• Calculate by comparing the student’s [test] scaled score with the state average scaled score and the student’s scaled score SEM o Below (-): state average scaled score – student’s scaled score SEM > student’s scaled Score o At (=): state average scaled score – student’s scaled score SEM <= student’s scaled Score <= state average scaled score + student’s scaled score SEM

Field	Description
	o Above (+): < student’s scaled Score > state average scaled score + student’s scaled score SEM
[Test] Item Score String	• Test Item Score/Response String • Apply Reporting Status test score rules formatting as appropriate for a specific deliverable • Each column in the string represents a core item (count’s toward student score) • Selected Response: o + = Correct Response o A,B,C,D = Incorrect Response o Z = No Response o X = Item Excluded from Student’s form • Writing Prompt: o 0,1,2 = Response Score o Z (blank) ,F (Foreign Language) ,P (Copy of Prompt) ,N (No Score) ,O (Off Topic) Non-Scorable Codes o X = Item Excluded from Student’s form
[Test] Field Item	• If at least one field test item is attempted on the test then “1”, otherwise “0”
ELA Reading Percent of Points Earned	• Percentage of possible points correct for reading items • Values: 0-100, N/A • Apply Reporting Status test score formatting rules as appropriate for a specific deliverable • Include all core items administered to the student
ELA Writing Percent of Points Earned	• Percentage of possible points correct for writing items • Values: 0-100, N/A • Apply Reporting Status test score rules as appropriate for a specific deliverable • Include all core items administered to the student
WR Trait Scores	• Student level writing trait scores are included part of overall ELA test • Apply Reporting Status test score formatting rules as appropriate for a specific deliverable • Refer to Writing Prompt: Dimension Score Adjustment table Student results column o 0,1,2 = Response Score o Z (blank) ,F (Foreign Language) ,P (Copy of Prompt) ,N (No Score) ,O (Off Topic) Non-Scorable Codes

Aggregate Data

Aggregation Level

Each student is assigned one State, District, and School code to use for aggregations as described in the table below

Aggregation Organizational Level	Aggregation Code
State	Partner Abbreviation
District	Combined Partner Abbreviation and District Code
School	Combined Partner Abbreviation, District and School Code

Aggregation Formulas

The aggregations below are calculated to support various datafiles and reports. The calculations are aggregated by state, school and district. *Student tests identified as Void/Duplicate or Remove are excluded from all aggregations.*

Aggregation	Calculation
Number Enrolled	• Number of student tests that have at least one test assigned one of the final reporting status values other than WDE or NLE for the aggregation level
Number Tested	• Number of student tests assigned TES, ESR, or IRR final reporting status for the aggregation level
Number of Did Not Test	• Number of student tests assigned ESM, INC, INV, PRF, ELL, EXE, DNT, WDR, NLE final reporting status for the aggregation level
Average Scale Score	• Average test scale score for students included in the “Number Tested” aggregation rounded to the nearest whole number for the aggregation level
Number of Students at each Performance Level	• Number of student tests included in the “Number Tested” count with the specific Performance Level Value for the aggregation level
Percent of Students at each Performance Level	• Divide the “Number of Students at each Performance Level” by the Number Tested for the aggregation level. Multiply by 100 and round to the nearest whole number.

Aggregation Suppression Rule

Aggregations with less than 10 students included in the denominator will be suppressed from state level reports only.

Post-Test Administration Data File Deliverables

Student Demographics Datafile (for Test Clean-up)

Description	• Cognia provides each participating Partner an excel file containing raw student data to support data cleanup
Generation Rules	• Refer to MSAA 2223 Student Demographic Instructions.pdf
File Layout	• Refer to MSAA 2223 Student Demographic Instructions.pdf
File Name	• Refer to MSAA 2223 Student Demographic Instructions.pdf

Organization Datafile (for Test Clean-up)

Description	• Cognia provides each participating Partner an excel file containing organization data to support data cleanup
Generation Rules	• ICORE organization data are used directly to create the file as detailed in the layout
File Layout	• MSAA2223OrgDataLayout.xlsx
File Name	• MSAA122_ICORE_[state abbreviation].xlsx

Test Materials Download Count Datafile

Description	• Breakthrough provides test materials data table • Cognia uses the data table to create a data file for each state containing the relevant state data
File Name	• MSAA2223_tblFilddownloads_[state abbreviation].xlsx

Writing Score Off-Topic Datafile

Description	• The writing off-topic datafile lists students and their writing prompt trait scores.
Generation Rules	• Raw ISCORE scores are provided in the file except Off Topic is O and B, F, N are translated to 0
File Layout	• State, DistrictCode, SchoolCode, DistrictName, SchoolName, Lname, Fname,StateStudentID, Grade, ItemNumber, Trait1score, Trait2Score, Trait3Score
File Name	• WritingDelivareble-[state abbreviaton].xlsx

Billable Records Datafile

Description	• MSAA States shall be billed out based on record results. Billable results shall be delivered to Cognia's Finance Department for true up and final billing.
Generation Rules	• Each tested student is considered a billable record o Each student test shall be considered a valid billable record when a test is launched and In Progress, Closed or Submitted. o A billable record does not include where a student does not have a test record, is no longer enrolled or is withdrawn • Records with a blank nap_delivery_id will be highlighted • The datafile will include two tabs: one for Reporting records (included in results datafile) and Not Reported Records (included in Duplicate/Void datafile) • The records will be reported in the file with their SSID
File Name	• Billing_[state abbreviation].xlsx

Scaled Score Lookup Datafile

Description	• The rawscore to scaled score lookup will be created and provided as an option to Partners
Generation Rules	• One EXCEL file for each Test Subject will be created containing the psychometric raw score to scale score lookup data • Each EXCEL file will contain a worksheet for each test grade
File Layout	• Each worksheet will contain columns: Grade, Subject, ScaleForm, RawScore, ScaledScore, LowScaledScore, HighScaledScore, and PerfLevel
File Name	• MSAA2223ScaledScoreLookups_mat.xlsx • MSAA2223ScaledScoreLookups_ela.xlsx • MSAA2223ScaledScoreLookups_sci.xlsx

Student Results School, District, State Datafile

Description	• The student results data file will contain all data for student tests not identified as Void/Duplicate during test Clean-Up as well as students tests added during test clean-up following the file layout • State files will be produced and provided on the sFTP • State, District, and School files will be provided to Breakthrough
Generation Rules	• The student results data file is sliced by state, district, and school. • Student tests are included in the specific version of the file based on the Aggregation Organization Level of State, District, and School assignment rules • Refer to table "Final Report Status Formatting of Student Scores table" • The file layout defines each field and valid values • The file will be exported to EXCEL.
File Layout	• MSAA2223StudentResultsLayout.xlsx • District and School files will contain a subset of variables as indicated in the layout "District, School Files" column • For Partners not participating in Science: worksheet StuResults_ELAMat will be used to generate the files • For Partners participating in Science: worksheet StuResults_ELAMATSCI will be used to generate the files • For 22-23, the files will be generated as follows • Pre-Standard Setting: All files posted to Breakthrough will follow the StuResultsELAMat layout. All files posted to the sFTP site for Partners who did not participate in Science will follow the StuResultsELAMat layout. All files posted to the sFTP site for Partners who did participate in Science, will follow the StuResults_ELAMATSCI, but Science scaled score and performance level data will be blank • Post-Standard Setting: All files (BT and sFTP) for Partners who participated in Science will receive updated files following StuResults_ELAMATSCI layout.
Preliminary State File Name	• 2023_[Partner abbreviation]_PreliminaryStudentResults.xlsx
State File Name	• 2023_[Partner abbreviation]_StateStudentResults.xlsx
BT State File Name	• 2023_[BT Org ID]_StateStudentResults.xlsx
BT District File Name	• 2023_[BT Org ID]_DistrictStudentResults.xlsx
BT School File Name	• 2023_[BT Org ID]_SchoolStudentResults.xlsx

Student Results Datafile: Final Reporting Status Formatting of Student Scores

Final Test Reporting Status	Code	State File All Scores*	District & Scaled Score	School File Perf Level	R/W Percent
Administration Irregularity	IRR	Yes	Yes	Yes	Yes: 0-100
Invalidated	INV	Yes	No	No	N/A
Parental Refusal	PRF	No	No	No	No
ELL Exempt (ELA Only)	ELL	No	No	No	N/A
Exempt	EXE	No	No	No	N/A
Withdrew	WDR	No	No	No	N/A
No Longer Eligible	NLE	No	No	No	N/A
Tested	TES	Yes	Yes	Yes	Yes: 0-100
Tested-Incomplete	INC	Yes	Yes	No	Yes: 0-100
Early Stopping Rule	ESR	Yes	Yes	Yes	N/A
Early Stopping Rule – Misadministration	ESM	Yes	Yes	No	Yes:0 -100
Did Not Test	DNT	No	No	No	N/A

(*) All Scores: State student results file includes item responses, WP trait scores, raw scores, scaled scores, and performance levels.

- Yes = Include score in data file; No = Leave column blank in data file; N/A = Put N/A in the data file

Duplicate/Void Student Datafile

Description	• The file contains the student tests identified as Void/Duplicate, including Off-Grade test records during test Clean-Up process • Data within the datafile shall be interpreted with caution since minimal Clean-Up has been applied
Generation Rules	• The file will follow the same layout and rules as the ELA/Math student results file, except only include student tests identified as Void/Dup • A file will be created for each Partner if there is at least one student test identified as Void/Dup • The file will be exported to EXCEL.
File Layout	• MSAA2223StudentResultsLayout.xlsx • District and School files will contain a subset of variables as indicated in the layout “District, School Files” column • For Partners not participating in Science: worksheet StuResults_ELAMat will be used to generate the files • For Partners participating in Science: worksheet StuResults_ELAMATSCI will be used to generate the files
State File Name	• 2023_[Partner abbreviation]_VoidDupResuls.xlsx

Post-Test Administration Report Deliverables

Student Report

Report Delivery

- Students who have an ELA or Math final reporting status of TES, ESR, or IRR will receive an ELA/Math Student Report.
- Students who have a Science final reporting status of TES, ESR, or IRR will receive a Science Student Report (Note: Science Student Report will be available after Standard Setting)

Print

- Only Partners who selected the Print option will receive two printed copies of the student report
- A print report package will be created by school.
- Slip sheets will be created at the start of each new report pack. The slip sheet identifies the appropriate shipping information and provides a way to track the secure shipment.
- ELA/Math Student Reports will be printed and shipped.
- Science Reports will be printed and shipped.
- Printed student reports will be gray-scaled.

Online

- A PDF will be generated for each Partner and school containing all student reports for the school regardless of test grade.
- Student reports will be sorted by Test Grade, Student Last Name, Student First Name, Student ID
- Prior to standard setting, the PDFs will only contain ELA/Math Student Reports. After Standard Setting, for Partners participating in Science, the online PDFs will be updated to add Science Student Reports at the end of each PDF. (Next year, the sort order can change)
- Online student reports will be in color.

Data Visualization

This section details the data visualizations for the ELA/Math and Science Student Report. Each ELA/Math student report is a two-page report (front and back). The ELA/Math report is designed to display both ELA and Math results side by side. The Science student report is a two-page report (front and back) one- subject report. The front page of every student report is noted as “Confidential”.

- Print Student First name possessive, when appropriate. Throughout the student report, the student’s first name appears embedded in text, it will appear as is or modified to be possessive as follows
  - If student first name ends in ‘s’ append apostrophe to student first name

- Otherwise, print [Student First Name]'s in section introduction sentence

- First Page Header
  - Name: [Student First Name] [Student Last Name]
  - ID: [State Student ID]
  - School: Print School Name
  - Test Date: Spring [Year] (example: Spring 2023)
  - Grade: if Test Grade 11 then High School else [Two-Digit Test Grade]
- First Page Performance Summary

Format Performance Summary section based on the student's final test reporting status as detailed in the table below.

Test Final Reporting Status	Test Result Section	Visualization
TES or IRR	**Performance Level**	• Print formatted earned student performance level o Level 1 o Level 2 o Level 3 o Level 4
	Score	• Print the student earned scaled score
	Score Graphic	• Place arrow in the relative location of the graphic for the student's scaled score with score printed above the arrow • Print scaled score ranges in each performance level
	Score Low/High	• Print the student's lower and upper scaled score
	R/W Percent of Points Earned (ELA Only)	• Print the student's earned percent of points
ESR	Performance Level	• Print formatted student performance with an asterisk • Level 1*
	Score	• Print the student scaled score provided by psychometrics. It is expected to be 1200.
	Score Graphic	• Place arrow in the relative location of the graphic for the student's scaled score with score printed above the arrow • Print scaled score ranges in each performance level
	Score Low/High	• Leave blank. Do not print the Low/High Scaled Score sentence.
	R/W Percent of Points Earned (ELA Only)	• Print N/A
All Other Values	**Leave blank under the Test Subject header except print the note:**	• Your child did not receive a score in this content area. Please contact your child's teacher/school for more information.

- First Page Performance Level Descriptors

Format Performance Level Descriptors section based on the student's final test reporting status detailed in the table below.

Test Final Reporting Status	Visualization
TES or IRR	• Print formatted performance level descriptors based on student test grade, test subject, and earned student performance level • The performance level descriptors were provided to Reporting during report design after standard setting. The text is carry forward from year to year. • Each statement starts with a checkmark
ESR	• Print the text under the Test Header: * Your child did not show an observable response mode during the test; therefore, the test was not administered by the teacher. If you have additional questions, please contact your child's teacher.
Other	• Leave section under Test header blank

- First Page Footer
  - Left Justified: Copyright information
  - Right justified: Page 1
- Second Page Header
  - 2023 Results for [Student First Name] [Student Last Name] ([State ID]) | "High School" or Grade [2-digit test grade] | [School Name]
  - Example: 2023 for Jane Smith (12345678) | Grade 04 | Demonstration School A
  - Example: 2023 for Jane Smith (12345678) | High School | Demonstration School A
- Second Page Letter to Parents and Guardians
  - Letter is provided by the Partner and one letter for all ELA/Math Student Reports
  - Letter is provided by the Partner and one letter for all Science Student Reports
- Second Page: What skills can be worked on next?

  Format "What skills can be worked on next?" section based on the student's final test reporting status as detailed in the table below.

Test Final Reporting Status	Visualization
TES or IRR	• Print the specific skills text provided during report design based on the students test grade and subject • Each statement starts with a plus symbol
ESR	• Print the text under the Test Header: • Revisit IEP communication goals in collaboration with the speech language pathologist, AT specialist, and others who assist the student in developing a consistent mode of communication.
Other	• Leave section under Test header blank

- Second Page What now?
  - Print the questions and suggestions developed during report design with student's first name embedded in the statements and questions

- Second Page Footer
    - Left Justified: Copyright information
    - Right justified: Page 1

School and District Roster Report

Report Delivery

Pre-Standard Setting:

- A School Roster Report will be produced when a school has at least one student assigned an ELA or Math reporting status value other than WDR or NLE.
- A District Roster Report will be produced when a district has at least one student assigned an ELA or Math reporting status value other than WDR or NLE.

Post-Standard Setting:

- A School Roster Report will be produced when a school has at least one student assigned an ELA, Math or Science reporting status value other than WDR or NLE.
- A District Roster Report will be produced when a district has at least one student assigned an ELA, Math or Science reporting status value other than WDR or NLE.

- Static PDFs will be generated to be posted online. The report is not printed.

Data Visualization

This section details the data visualizations for the School and District Roster Report.

- **District Roster Report**
    - Header
        - Print: CONFIDENTIAL
        - Print: [Formatted State Name]
        - Print: [Formatted District Name]
        - Print: If test grade =11 then High School else Grade [Two Digit Test Grade]
    - Summary Data Rows:
        - Each row will contain the state and district aggregated test results
        - Do not suppress aggregations
    - Student Roster
        - Header: Spring 2023
        - Student Name [Student Last Name],[Student First Name]
        - Student ID [State Student ID]
        - Test Status Impact on Report of Student Test Results

Final Test Reporting Status	Code	Print Test Status	Print State Compare	Print Scale Score	Print Performance Level
Administration Irregularity	IRR	Yes	Yes	Yes	Yes
Invalidated	INV	Yes	No	No	No
Parental Refusal	PRF	Yes	No	No	No
ELL Exempt (ELA Only)	ELL	Yes	No	No	No
Exempt	EXE	Yes	No	No	No
Withdrew	WDR	Yes	No	No	No
No Longer Eligible	NLE	Yes	No	No	No
Tested	TES	No	Yes	Yes	Yes
Tested-Incomplete	INC	Yes	Yes	Yes	No
Early Stopping Rule	ESR	Yes	Yes	Yes	Yes
Early Stopping Rule – Misadministration	ESM	Yes	Yes	Yes	No
Did Not Test	DNT	Yes	No	No	No

Print Test Status: Yes – print the three-letter code; No – Leave blank

Print State Compare: Yes – print - , +, or = based on student score; No – Leave blank

Print Scale Score: Yes – print student scale score; No – Leave blank

Print Performance Level: Yes: Print "Level 1", "Level 2" , "Level 3", or "Level 4" student performance level; No – Leave blank

- Footer
  - State Comparison Key
  - Copyright
  - Page X (Restart page count at 1 for each test grade)

School, District, and State Summary Report

Report Delivery

- Each participating Partner with at least one student included the "Number Enrolled" calculation will receive a State Summary Report.
- Each district with at least one student included the "Number Enrolled" calculation will receive a District Summary Report.
- Each school with at least one student included in the "Number Enrolled" will receive a School Summary Report.
- Static PDFs will be generated to be posted online. The report is not printed.
- Pre-Standard Setting: Each static PDF will contain a page for ELA and a page for mathematics
- Post-Standard Setting: For Partners who participated in science, each static PDF will contain a page for ELA, a page Math, and a page for science

Data Visualization

This section details the data visualizations for the State, District, and School Summary Report

- **State Summary Report**
  - Title: [Formatted Subject]
  - Right Justified Header: Print [Formatted State Name]
  - Summary Data Rows:
    - Each row will contain the state aggregated test results for each grade
    - If the “Number Tested” is less than 10, then suppress the Number and Percent at each Performance Level and Average Scale Score
  - Footnote: Copyright statement
- **District Summary Report**
  - Title:
    - Print CONFIDENTIAL
    - Print [Formatted Subject]
  - Right Justified Header:
    - Print [Formatted State Name]
    - Print [District Name]
  - Summary Data Rows:
    - Each grade row will contain the state and district aggregated test results
    - Only grades with at least one student enrolled in the district will be included on the district roster
    - Do not suppress aggregations
  - Footnote: Copyright statement
- **School Summary Report**
  - Title:
    - Print: CONFIDENTIAL
    - Print: [Formatted Subject]
  - Right Justified Header:
    - Print: [Formatted State Name]
    - Print: [District Name]
    - Print: [School Name]
  - Summary Data Rows:
    - Each grade row will contain the state, district, and school aggregated test results
    - Only grades with at least one student enrolled in the school will be included on the district roster
    - Do not suppress aggregations
  - Footnote: Copyright statement

eMetric Data Interaction

Student & Summary Results	
Description	• Cognia will provide eMetric data to support eMetric Data Interaction reporting for Partners who opted into this option • eMetric will receive two types of files: Student Results, Summary Results
Generation Rules	• Exclude Void/Duplicate Student Tests from the student results file (and subsequently aggregations in summary files) • Pre-Standard Setting: All Science specific fields will be blank in the student results files; cience summary data rows will be excluded. • The file layouts define each field and valid values • The student files will be exported to EXCEL • The summary files will be exported to EXCEL.
File Layouts	• Student Results: MSAA2223StudentResultsLayout.xlsx worksheet StuResults_ELAMATSCI • Summary: MSAA2223eMetricSummaryDataTransfer.xlsx
File Names	• MSAA2223_ [Partner abbreviation]_StudentResults.xlsx • MSAA2223_ [Partner abbreviation]_[test grade]SummaryData.xlsx

PDF Metadata	
Description	• Cognia will provide eMetric data to support the eMetric PDF download hub for Partners who opted into this option
Generation Rules	• Each school student report PDF will be included in the CSV • The first row will contain field names: ProgramName, ReportName, Org_Num, PDF_Name
File Layouts	• Program Name: MSAA (Alternate Assessment) • Year: 2023 • ReportName: Individual Student Report • Org_Num: <client district code>-<client school code> • PDF_Name: <school student report PDF name>
File Names	• MSAA2223_ [Partner abbreviation]_eMetricPDFMetaData.csv

Focal Point Reporting

Student & Summary Results	
Description	• Cognia will provide Focal Point data to support Focal Point reporting for Partners who opted into this option • Focal Point will receive two types of files: Student Results, Summary Results
Generation Rules	• Exclude Void/Duplicate Student Tests from the student results file (and subsequently aggregations in summary files) • Pre-Standard Setting: All Science specific fields will be blank in the student results files; science summary data rows will be excluded. • The file layouts define each field and valid values • The student files will be exported to EXCEL • The summary files will be exported to EXCEL.
File Layouts	• Student Results: MSAA2223StudentResultsLayout.xlsx worksheet StuResults_ELAMATSCI • Summary: MSAA2223eMetricSummaryDataTransfer.xlsx
File Names	• MSAA2223_ [Partner abbreviation]_StudentResults.xlsx • MSAA2223_ [Partner abbreviation]_[test grade]SummaryData.xlsx

Parental Rescore Request

For Partners selecting the Parental Rescore Request option, if one or more students require a score update as part of the parental rescore request the following deliverables will be updated with the corrected student scores and provided to Client Services Program Management to be delivered to each Partner. Aggregate data will not be re-calculated as part of the parental rescore request.

- Student Results Datafile
- Student Report
- School and District Roster Report

APPENDIX H
2023 GUIDE FOR SCORE REPORT INTERPRETATION

Multi-State Alternate Assessment

2023 Guide for Score Report Interpretation

MSAA Service Center

Phone: (866) 834-8879
Email: MSAAServiceCenter@cognia.org
MSAA Online Assessment System: www.msaaassessment.org

State-Specific Information

Listed below is the contact information for each state's MSAA State Representative(s):

American Samoa	Arizona	Bureau of Indian Education
Thor Tinitali 684-633-1323 ext. 226 thort@doe.as Kim Pilitati 684-633-4789 ext. 238 kim.pilitati@doe.as	Bethany Spangenberg 602-542-4061 Sarah Han 602-364-0452 AlternateAssessment@azed.gov	Donald Griffin 703-282-3316 Donald.Griffin@bie.edu Aurelia Shorty 505-274-3746 Aurelia.Shorty@bie.edu
CNMI	**District of Columbia**	**DoDEA**
Fasefulu Tigilau 670-789-8739 Fasefulu.Tigilau@cnmipss.org June De Leon 671-735-2481 June.DeLeon@guamcedders.org	Stephanie Snyder 202-765-7158 Stephanie.Snyder@dc.gov Asaad Fulton 202-899-6141 Asaad.Fulton@dc.gov	Dr. Blessing Mupanduki 571-372-7983 blessing.mupanduki@dodea.edu Jaclyn Haynes 571-372-6008 jaclyn.haynes@dodea.edu
Guam	**Maine**	**Montana**
Terese Crisostomo 671-300-1323 tdcrisostomo@gdoe.net June De Leon 671-735-2481 June.DeLeon@guamcedders.org	Jodi Bossio-Smith 207-530-1462 jodi.bossio-smith@maine.gov	Austin Waldbillig 406-444-0748 Austin.Waldbillig@mt.gov Assessment Help Desk 844-867-2569 OPIAssessmentHelpDesk@mt.gov
South Dakota	**Tennessee**	**United States Virgin Islands**
Stacy Holzbauer 605-295-3441 Stacy.Holzbauer@state.sd.us Chris Booth 605-773-6156 Christina.Booth@state.sd.us	For teachers, contact your district test coordinator (TC). For Scoring & Accountability questions, contact: TNED.Accountability@tn.gov For district TCs, contact: Nancy Williams Nancy.E.Williams@tn.gov	Alexandria Baltimore-Hookfin 340-773-1095 ext.7084 Alexandria.Baltimore@vide.vi
Vermont		
Please contact the Agency of Education at: AOE.SpecialEd@vermont.gov		

Table of Contents

Introduction to the MSAA

Purpose

The Multi-State Alternate Assessment (MSAA) is a comprehensive assessment system, designed to promote increasing higher academic outcomes for students with the most significant cognitive disabilities, in preparation for a broader array of post-secondary outcomes. The MSAA is designed to assess students with the most significant cognitive disabilities and measures academic content that is aligned to and derived from each participating state's content standards. This assessment contains many built-in supports that allow students to use materials they are most familiar with, and communicate what they know and can do as independently as possible. The MSAA is administered in the areas of English Language Arts (ELA) and Mathematics in grades 3–8 and high school. American Samoa, Arizona, the Bureau of Indian Education (BIE), Commonwealth of the Northern Mariana Islands (CNMI), Guam, Maine, the United States Virgin Islands (USVI), and Vermont also administered Science in grades 5, 8, and high school.

This assessment was developed with Cognia through the research and development done by the National Center and State Collaborative (NCSC), and is now carried forward by the MSAA Partners, including American Samoa, Arizona, BIE, CNMI, Department of Defense Education Activity (DoDEA), District of Columbia, Guam, Maine, Montana, South Dakota, Tennessee, USVI, and Vermont.

This guide provides information regarding the administration and results of the spring 2023 MSAA to district and school personnel.

Student Participation

The criteria for student participation in the MSAA reflect the pervasive nature of a significant cognitive disability. All content areas should be considered by the Individualized Education Program (IEP) team when determining who should participate in this assessment. The table below shows the participation criteria and the descriptors used to determine eligibility for participation for each student. Students must meet the following eligibility criteria:

Participation Criteria	Participation Criteria Descriptors
1. The student has a significant cognitive disability.	Review of student records indicates a disability or multiple disabilities that significantly impact intellectual functioning and adaptive behavior.* *Adaptive behavior is defined as essential for someone to live independently and to function safely in daily life.
2. The student is learning content linked to grade-level content standards.	Goals and instruction listed in the IEP for this student are linked to the enrolled grade-level content standards and address knowledge and skills that are appropriate and challenging for this student.
3. The student requires extensive direct individualized instruction and substantial supports to achieve measurable gains in the grade and age-appropriate curriculum.	The student (a) requires extensive, repeated, individualized instruction and support that is not of a temporary or transient nature, and (b) uses substantially adapted materials and individualized methods of accessing information in alternative ways to acquire, maintain, generalize, demonstrate, and transfer skills across multiple settings.

Assessments for students with the most significant cognitive disabilities rely on a foundation of communicative competence. Students who do not have receptive and expressive communication are unlikely to be able to demonstrate what they know and can do on an assessment. Students who do not have a mode of communication are identified during the assessment process.

Post assessment, teachers may use the Communication Toolkit developed by NCSC to help these students develop a mode of communication. The Toolkit can be found here: wiki.ncscpartners.org/index.php/Communication_Tool_Kit.

Overview of the MSAA Format

The MSAA assesses ELA (reading and writing) and mathematics at grades 3–8 and high school and is aligned to the state's content standards and the MSAA Core Content Connectors. The MSAA is a computer-based, on-demand, stage-adaptive assessment consisting mostly of selected response and some constructed-response items written at three levels of complexity. These complexity levels represent different levels of skill acquisition by students.

Students with the most significant cognitive disabilities often need materials and instructional strategies that are substantially adapted, scaffolded, and have built-in supports to meet their individual needs.

The MSAA levels of complexity are designed to follow instructional practices. When students begin to learn a new skill, or acquire new knowledge, they need more support. As students learn and develop mastery of that skill or knowledge, they need less support. The test items on the MSAA are developed with many scaffolds and supports embedded within the items. Supports not embedded in the test items may be provided as accommodations, as well as other allowable ways to present the item to a student, based on their individual requirements.

The assessment is a computer-based test and is administered one-on-one. Based on the needs of the student, the assessment may also be delivered in a paper-pencil format. The needs of the student may also be addressed through other supports and accommodations, such as reading the test aloud, having a scribe, using manipulatives, using object replacement, translating the test into American Sign Language, among others. Test administrators (TAs) have substantial leeway in developing a testing schedule, with the ability to start and stop a test depending on the engagement of the student.

Each content area consists of 45–55 items across two test sessions. These are primarily selected-response items with some constructed-response items. The writing portion of the ELA test contains a scaffolded writing prompt at each grade level.

American Samoa, Arizona, BIE, CNMI, Guam, Maine, USVI, and Vermont also administered Science in grades 5, 8, and high school. The Science test is aligned to the state's content standards and the Extended Performance Expectations. The science assessment is also computer-based and consists of selected-response items.

Scoring

Scoring of most items is accomplished within the online test platform. The selected-response items are scored as correct or incorrect by the test platform based on the answer keys programmed into the system. Constructed-response items are scored by the TA and then marked correct or incorrect in the test platform. Items without responses receive a score of zero. Student responses to writing prompts are hand scored by trained scorers utilizing the rubrics in Appendix A.

MSAA Score Reports

Overview

This guide describes the types of score reports provided for the 2022–23 MSAA administration. The data in the sample reports are for illustrative purposes only and are not intended to reflect performance of any student(s).

Information included on the score reports:

- ***Performance Levels*** describe how the student performed in relation to the knowledge and skills of that content area and grade level. Each performance level has two components: the scale scores that make up each level and the performance level descriptors (PLD). The PLDs are broad and general statements regarding skills and abilities of students who have attained each level.
    - Performance levels for ELA and mathematics for the MSAA were established by committees of educators after the first NCSC administration of the assessment in 2015 and were updated in 2018. PLDs for each grade level of ELA and mathematics can be found in Appendix B. The scale score ranges that make up each performance level for ELA and mathematics can be found in Appendix C.
    - Content and Accessibility specialists collaborated with MSAA Science Partners to develop PLDs for science in 2022. Science PLDs consist of policy PLDs and range PLDs. Policy PLDs provide high-level student performance expectations, and range PLDs describe the knowledge, skills, and abilities that students must demonstrate to be classified into a performance level. PLDs for grades 5, 8, and high school science can be found in Appendix B.
- ***Scale scores*** report the performance level the student achieved. Scale scores are more precise than performance levels and may be used to make comparisons between groups of students, schools, and districts. In Appendix C, Table 1 shows the scale score ranges for each performance level and grade level for ELA and mathematics.
- ***Descriptive and informative reports***. In addition to including student demographic information, performance level, and scale scores, the Individual Student Report (ISR) contains supportive information about student performance and MSAA measures.
    - ***Reading and Writing Scores***—the percentage of items answered correctly for reading and writing separately. The writing items consisted of selected response and constructed response (or multiple choice and the writing prompt).
    - ***What skills can be worked on next?***—skills related to the standards in the following grade.
    - ***What now?***—conversation starters for parents when talking with teachers about instruction for their child.

Interpreting and Using the MSAA Scores

The MSAA tests student performance based on the state's content standards at the student's enrolled grade level. The student's performance level is based on alternate academic achievement standards. Results for the MSAA are reported by a scale score and performance level for each content area.

MSAA scores should be used in conjunction with the IEP progress reports, student work, diagnostic assessments, district-required assessments, and report cards in order to place the student's performance on academic content and skills in context and to provide a complete picture of the student's progress across a wide range of categories.

It is helpful to read the PLDs to understand the expectations for the performance level and grade level for each student. This information can provide a concrete link from the test to instructional planning.

Talking to Parents and Guardians

MSAA parent overviews are available for parents to introduce and describe the assessment. To view the parent guides, visit www.msaastates.com and select the "Resources for Families" tab. You may also contact your MSAA State Representative to locate these materials.

When talking to parents and guardians about their child's score, it may be helpful to keep the following in mind:

- MSAA assessment results should be used along with local assessment results and other information to determine what changes in curriculum and instruction may be needed to support their student's learning.
- MSAA scores alone should not be used to make placement or eligibility decisions.

Special Reporting Codes and Messages

In some cases, students were assigned a special reporting code. A complete list of special reporting codes and their associated descriptions is provided below. For additional information or interpretation of special reporting codes, contact your MSAA State Representative.

Code	Test Status	Description
ESR	Early Stopping Rule	If the TA did not observe a student response after the presentation of four items, the test was closed by the test coordinator (TC).
ESM	Early Stopping Rule Misadministration	Testing may have ended early on the basis that a consistent mode of communication was not observed. At least one response was recorded for the student, but the student may not have had the opportunity to complete the entire test.
INC	Tested – Incomplete	The student's test was not submitted by the close of testing. The student may not have had the opportunity to complete the entire test.
TES	Test	The student's test was submitted by the close of testing.
IRR	Administration Irregularity	An administration irregularity not necessitating an invalidation of scores was reported for the student's test.
INV	Invalidated	The results of the student's test have been invalidated.
PRF	Parental Refusal	The student did not test due to a parent/guardian refusal.
ELL	ELL Exempt (ELA Only)	The student was exempt from ELA testing due to being a first year English Language Learner.
EXE	Exempt (Emergency, Medical, Other)	The student was exempt from testing.
DNT	Did Not Test	The student did not test via the MSAA assessment.
WDR	Withdrew	The student withdrew.
NLE	No Longer Eligible	The student is not eligible to test via the MSAA assessment.

Types of Score Reports

Below are the types of MSAA score reports that will be available on the MSAA Reporting Portal. Only district TCs using their current MSAA username and password may access the MSAA reports here: www.msaaassessment.org under the Reporting tab. Reports are only available during the online reporting window. All MSAA score reports are confidential documents.

- Reports for the District
  - District Summary Report (DSR)
  - District Roster Report (DRR)
  - Student Results File
- Reports for the School
  - School Summary Report (SSR)
  - School Roster Report (SRR)
  - Individual Student Report
  - Student Results File

An Excel file of all student results at the district and school level will be available to district TCs through the MSAA Reporting Portal. For information regarding this file or questions about accessing the reports, contact your MSAA State Representative. Contact information can be found at the beginning of this document.

Testing Participation

All students in grades 3–8 and high school are required to be assessed in ELA and mathematics. Participation status is assigned independently for ELA and mathematics.

All submitted tests receive a participation status, regardless of the number of item responses.

For additional information regarding the reported test status, contact your MSAA State Representative. Contact information can be found at the beginning of this document.

District Summary Report

The DSR provides district staff with a summary of student participation and performance by district and school. State-level data is taken from the individual participating state. See Figure 1 below.

Figure 1. Sample District Summary Report

CONFIDENTIAL

1 English Language Arts

2 **SUMMARY REPORT**
Demonstration State
Demonstration District B

		Enrolled	Tested (3)	Did Not Test	Average Scale Score	Performance Level (4) Level 1 N	Level 1 %	Level 2 N	Level 2 %	Level 3 N	Level 3 %	Level 4 N	Level 4 %
03	State	22	12	10	1225	5	42	6	50	1	8	0	0
	District	22	12	10	1225	5	42	6	50	1	8	0	0
04	State	17	9	8	1219	7	78	1	11	1	11	0	0
	District	17	9	8	1219	7	78	1	11	1	11	0	0
05	State	28	14	14	1223	8	57	3	21	2	14	1	7
	District	28	14	14	1223	8	57	3	21	2	14	1	7
06	State	17	10	7	1222	5	50	2	20	3	30	0	0
	District	17	10	7	1222	5	50	2	20	3	30	0	0
07	State	17	10	7	1223	5	50	4	40	1	10	0	0
	District	17	10	7	1223	5	50	4	40	1	10	0	0
08	State	21	12	9	1221	5	42	3	25	4	33	0	0
	District	21	12	9	1221	5	42	3	25	4	33	0	0
High School	State	26	15	11	1223	8	53	3	20	4	27	0	0
	District	26	15	11	1223	8	53	3	20	4	27	0	0

The DSR contains the following features, highlighted above:

1. Content area of the report.
2. State and district included in the report.
3. Number of students by grade who were enrolled, tested, did not test, and average scale score by state and district.
4. The number and percentage of students at each performance level by grade in the state and district.

District Roster Report

The DRR provides district staff with a summary of student scale scores and performance levels by district and state. State-level data is taken from the individual participating state. See Figure 2 below.

Figure 2. Sample District Roster Report

CONFIDENTIAL

Multi-State Alternate Assessment

2 DISTRICT ROSTER REPORT
Demonstration State
Demonstration District B

Grade 03

1 Mathematics

	Enrolled	Tested	Average Scale Score	Level 1 (%)	Level 2 (%)	Level 3 (%)	Level 4 (%)
State	22	13	1234	31	0	62	8
District	22	13	1234	31	0	62	8

3

Spring 2023

Student Name Student ID	Mathematics			
	Test Status*	State Compare	Scale Score	Performance Level
LASTNAME56, FIRST56 D056	DNT			
LASTNAME67, FIRST67 D067	ESR	-	1200	Level 1
LASTNAME77, FIRST77 D077	ESR	-	1200	Level 1
LASTNAME79, FIRST79 D079	ESR	-	1200	Level 1
LASTNAME81, FIRST81 D081	ESR	-	1200	Level 1
LASTNAME83, FIRST83 D083		+	1245	Level 3
LASTNAME91, FIRST91 D091		+	1253	Level 3

4

State Comparison Key
- \- Performance is lower than state average
- = Performance is similar to state average
- \+ Performance is greater than state average

* For descriptions of the Test Statuses, see your State's Guide for Score Report Interpretation.

 Page 2

The DRR contains the following features, highlighted above:

1. Content area of the report.
2. State and district included in the report.
3. Number of students who were enrolled, tested, the average scale score, and the percentage of students at each performance level by state and district.
4. The test status, state comparison, scale score, and performance level by student and content area. Refer to the Special Reporting Codes and Messages for information regarding test status.

School Summary Report

The SSR provides summarized performance information at the state, district, and school level for each grade, including number of students enrolled, tested, did not test, as well as average scale score and performance level. See Figure 3 below.

Figure 3. Sample School Summary Report

msaa
Multi-State Alternate Assessment

CONFIDENTIAL

1 English Language Arts

2 **SUMMARY REPORT**
Demonstration State
Demonstration District B
Demonstration School 4

Grade		Enrolled	3 Tested	Did Not Test	Average Scale Score	4 Performance Level Level 1 N	Level 1 %	Level 2 N	Level 2 %	Level 3 N	Level 3 %	Level 4 N	Level 4 %
03	State	22	12	10	1225	5	42	6	50	1	8	0	0
	District	22	12	10	1225	5	42	6	50	1	8	0	0
	School	22	12	10	1225	5	42	6	50	1	8	0	0
04	State	17	9	8	1219	7	78	1	11	1	11	0	0
	District	17	9	8	1219	7	78	1	11	1	11	0	0
	School	17	9	8	1219	7	78	1	11	1	11	0	0
05	State	28	14	14	1223	8	57	3	21	2	14	1	7
	District	28	14	14	1223	8	57	3	21	2	14	1	7
	School	28	14	14	1223	8	57	3	21	2	14	1	7
06	State	17	10	7	1222	5	50	2	20	3	30	0	0
	District	17	10	7	1222	5	50	2	20	3	30	0	0
	School	17	10	7	1222	5	50	2	20	3	30	0	0
07	State	17	10	7	1223	5	50	4	40	1	10	0	0
	District	17	10	7	1223	5	50	4	40	1	10	0	0
	School	17	10	7	1223	5	50	4	40	1	10	0	0
08	State	21	12	9	1221	5	42	3	25	4	33	0	0
	District	21	12	9	1221	5	42	3	25	4	33	0	0
	School	21	12	9	1221	5	42	3	25	4	33	0	0
High School	State	26	15	11	1223	8	53	3	20	4	27	0	0
	District	26	15	11	1223	8	53	3	20	4	27	0	0
	School	26	15	11	1223	8	53	3	20	4	27	0	0

The SSR contains the following features, highlighted above:

1. Content area of the report.
2. State, district, and school included in the report.
3. Number of students by grade who were enrolled, tested, did not test, and average scale score by state, district, and school.
4. The number and percentage of students at each performance level by grade in the state, district, and school.

School Roster Report

The SRR provides student performance information at the school level for each grade, including each student's test status, scale score, and performance level. See Figure 4 below.

Figure 4. Sample School Roster Report

msaa
Multi-State Alternate Assessment

CONFIDENTIAL

SCHOOL ROSTER REPORT

1 Demonstration State
Demonstration District B
Demonstration School 4
Grade 05

English Language Arts

2	Enrolled	Tested	Average Scale Score	Level 1 (%)	Level 2 (%)	Level 3 (%)	Level 4 (%)
State	28	14	1223	57	21	14	7
District	28	14	1223	57	21	14	7
School	28	14	1223	57	21	14	7

Spring 2023

	3 English Language Arts			
Student Name Student ID	Test Status*	State Compare	Scale Score	Performance Level
4 LASTNAME10, FIRST10 D010		+	1230	Level 1
LASTNAME11, FIRST11 D011	ESR	-	1200	Level 1
LASTNAME12, FIRST12 D012	ESR	-	1200	Level 1
LASTNAME133, FIRST133 D133		+	1240	Level 3
LASTNAME137, FIRST137 D137	ESM	-	1200	
LASTNAME141, FIRST141 D141	DNT			
LASTNAME144, FIRST144 D144	DNT			
LASTNAME148, FIRST148 D148	ESM	-	1200	
LASTNAME149, FIRST149 D149	ESM	-	1200	
LASTNAME152, FIRST152 D152	DNT			
LASTNAME155, FIRST155 D155	DNT			
LASTNAME157, FIRST157 D157	DNT			
LASTNAME160, FIRST160 D160	INV			
LASTNAME161, FIRST161 D161	INV			
LASTNAME21, FIRST21 D021	ESM	-	1200	

5 **State Comparison Key**
- \- Performance is lower than state average
- = Performance is similar to state average
- \+ Performance is greater than state average

* For descriptions of the Test Statuses, see your State's Guide for Score Report Interpretation.

 Page 1

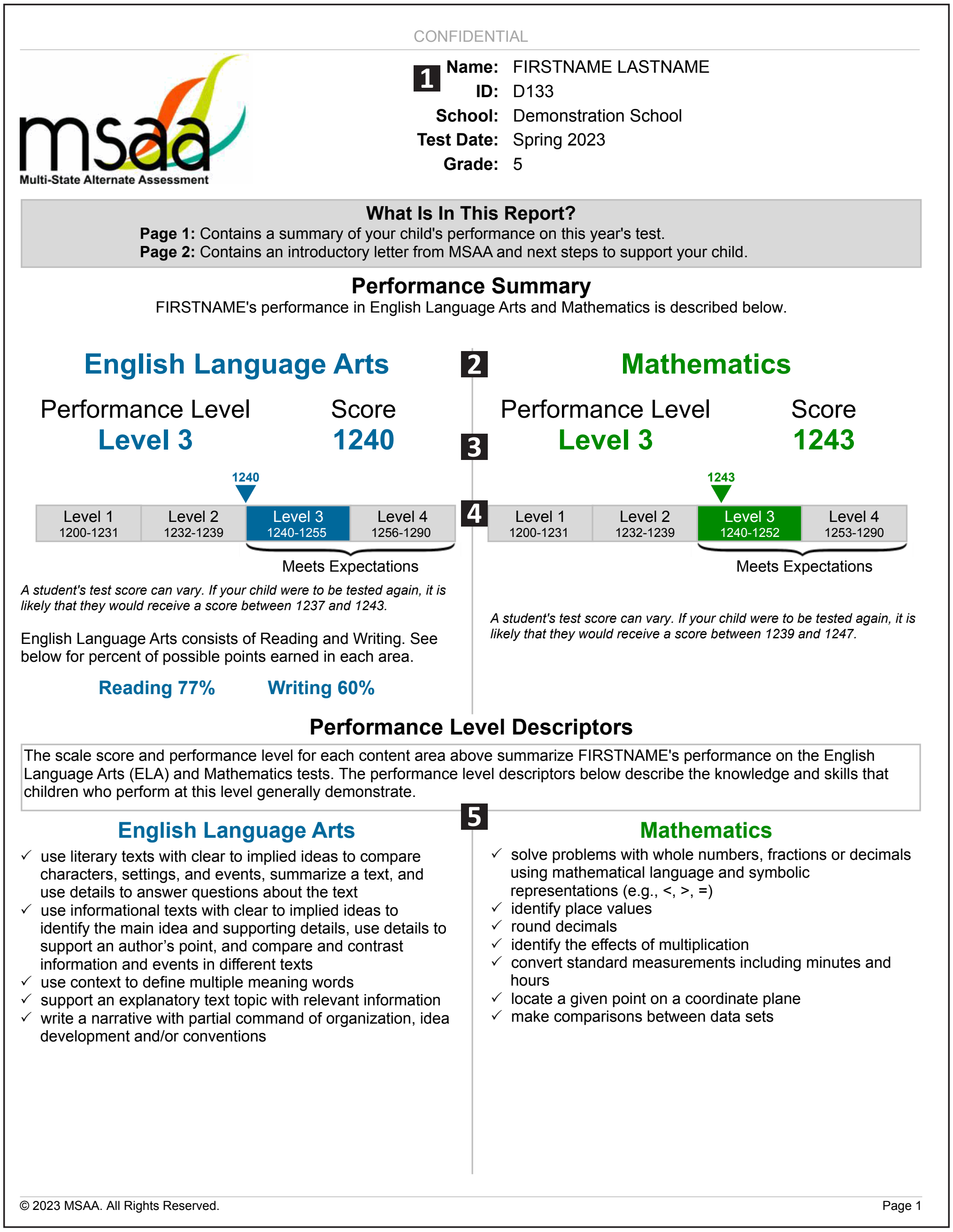
The SRR contains the following features, highlighted above:

1. The state, district, and school included in the report.
2. A summary of enrolled and tested students and the average scale score for the state, district, and reported school. The results are displayed by content area.
3. For each content area, the student's test status, comparison to other students in the same grade level in the state, scale score, and performance level are displayed.
4. This section of the report includes all students tested at the school for the specified grade.
5. This key shows symbols used in the "State Compare" column.

Individual Student Report

The ISR provides scale score and performance level information for a specific student. Figure 5 shows page 1 of the ISR. Full samples of the ISR are included in Appendix D.

Figure 5. Sample Individual Student Report



CONFIDENTIAL

msaa Multi-State Alternate Assessment

1 **Name:** FIRSTNAME LASTNAME
ID: D133
School: Demonstration School
Test Date: Spring 2023
Grade: 5

What Is In This Report?
Page 1: Contains a summary of your child's performance on this year's test.
Page 2: Contains an introductory letter from MSAA and next steps to support your child.

Performance Summary

FIRSTNAME's performance in English Language Arts and Mathematics is described below.

2

English Language Arts

Performance Level: **Level 3** | Score: **1240**

3

1240

4

Level 1	Level 2	Level 3	Level 4
1200-1231	1232-1239	1240-1255	1256-1290

Meets Expectations

A student's test score can vary. If your child were to be tested again, it is likely that they would receive a score between 1237 and 1243.

English Language Arts consists of Reading and Writing. See below for percent of possible points earned in each area.

Reading 77% **Writing 60%**

Mathematics

Performance Level: **Level 3** | Score: **1243**

1243

Level 1	Level 2	Level 3	Level 4
1200-1231	1232-1239	1240-1252	1253-1290

Meets Expectations

A student's test score can vary. If your child were to be tested again, it is likely that they would receive a score between 1239 and 1247.

Performance Level Descriptors

The scale score and performance level for each content area above summarize FIRSTNAME's performance on the English Language Arts (ELA) and Mathematics tests. The performance level descriptors below describe the knowledge and skills that children who perform at this level generally demonstrate.

5

English Language Arts

- ✓ use literary texts with clear to implied ideas to compare characters, settings, and events, summarize a text, and use details to answer questions about the text
- ✓ use informational texts with clear to implied ideas to identify the main idea and supporting details, use details to support an author's point, and compare and contrast information and events in different texts
- ✓ use context to define multiple meaning words
- ✓ support an explanatory text topic with relevant information
- ✓ write a narrative with partial command of organization, idea development and/or conventions

Mathematics

- ✓ solve problems with whole numbers, fractions or decimals using mathematical language and symbolic representations (e.g., <, >, =)
- ✓ identify place values
- ✓ round decimals
- ✓ identify the effects of multiplication
- ✓ convert standard measurements including minutes and hours
- ✓ locate a given point on a coordinate plane
- ✓ make comparisons between data sets

Page 1

The ISR contains the following features, highlighted above:

1. The report header includes the student's full name, student ID, school, and grade.
2. The results for each content area are displayed separately on the report.
3. The student's scale score and performance level for each content area are shown.
4. This display shows the student's score compared to the performance level scale.
5. This text shows the PLD for the student's performance level.

Grade 3 Writing Scoring Rubric

Level 2

Rubric Elements	Full Evidence 3	Partial Evidence 2	Limited Evidence 1	Unrelated Evidence 0 or 5	
Organization – The narrative establishes a situation (activity and setting) and includes a character with relevant descriptive statements. The response provides a conclusion.	The narrative includes at a minimum: • character **and** situation (activity **and** setting) • a conclusion that follows from the narrated experiences **or** events	The narrative includes at a minimum: • character **and** situation (activity **or** setting) • a conclusion that **may not** follow from the narrated experiences **or** events	The narrative includes at a minimum: • **some** evidence related to a character, situation (activity **or** setting), **or** conclusion	**0** • **no** evidence of organization	**5** • evidence is **off topic**
Idea Development – The narrative includes a sequence of events that unfold naturally and develops a story using temporal words.	The narrative includes at a minimum: • a sequence of **two** events related to the situation (activity **or** setting) • **both** events include a detail	The narrative includes at a minimum: • **two** events related to the situation (activity **or** setting) • **one** of the events includes a detail	The narrative includes at a minimum: • **one** event related to the situation (activity **or** setting)	**0** • **no** evidence of idea development	**5** • evidence is **off topic**
Conventions – Students use standard English conventions (e.g., end punctuation, subject-verb agreement).	The narrative includes **more than one sentence and** at a minimum: • end punctuation for **more than one** thought unit • **one** simple sentence that contains a complete thought **with** subject-verb agreement (e.g., “Dog runs” or “dog runs”)	The narrative includes at a minimum: • end punctuation for **one** thought unit • **one** thought unit **with or without** subject-verb agreement	The narrative includes at a minimum: • **one** use of standard English conventions (end punctuation for **one** thought unit **or one** thought unit **with or without** subject-verb agreement)	**0** • **no** evidence of standard English conventions	

Grade 3 Writing Scoring Rubric

Level 3

Rubric Elements	Full Evidence 3	Partial Evidence 2	Limited Evidence 1	Unrelated Evidence 0 or 5	
Organization – The narrative establishes a situation (activity and setting) and includes a character with relevant descriptive statements. The response provides a conclusion.	The narrative includes at a minimum: • character **and** situation (activity **and** setting) • **two** descriptions related to a character • a conclusion that follows from the narrated experiences **or** events	The narrative includes at a minimum: • character **and** situation (activity **or** setting) • **one** description related to a character • a conclusion that **may not** follow from the narrated experiences **or** events	The narrative includes at a minimum: • **some** evidence related to a character, situation (activity **or** setting), **or** conclusion **OR** • descriptive words related to a character **or** situation (activity **or** setting)	**0** • **no** evidence of organization	**5** • evidence is **off topic**
Idea Development – The narrative includes a sequence of events that unfold naturally and develops the story using temporal words (e.g., first, then, next).	The narrative includes at a minimum: • **two** sequenced events related to the situation (activity **or** setting) • **both** events include a detail • appropriate use of temporal words that signal order of events	The narrative includes at a minimum: • **two** events related to the situation (activity **or** setting) • **one** of the events includes a detail • **one** temporal word that may **or** may not be used appropriately	The narrative includes at a minimum: • **one** event related to the situation (activity **or** setting)	**0** • **no** evidence of idea development	**5** • evidence is **off topic**
Conventions – Students use standard English conventions (e.g., capitalization, end punctuation, subject-verb agreement).	The narrative includes **more than one sentence and** at a minimum: • capitalization at the beginning of the **majority** of thought units • end punctuation for **more than one** thought unit • **one** simple sentence that contains a complete thought **with** subject-verb agreement (e.g., "Dog runs" or "dog runs")	The narrative includes at a minimum **two** of the following: • capitalization at the beginning of **one** thought unit • end punctuation for **one** thought unit • **one** simple sentence **with or without** subject-verb agreement	The narrative includes at a minimum: • **one** use of standard English conventions (capitalization at the beginning of **one** thought unit, end punctuation for **one** thought unit **or one** thought unit **with or without** subject-verb agreement)	**0** • **no** evidence of standard English conventions	

Grade 4 Writing Scoring Rubric

Level 2

Rubric Elements	Full Evidence 3	Partial Evidence 2	Limited Evidence 1	Unrelated Evidence 0 or 5	
Organization – The narrative establishes a situation (activity **or** setting) and includes a character. The response provides a conclusion.	The narrative includes at a minimum: • character **and** situation (activity **or** setting) • a conclusion that follows from the narrated experiences **or** events	The narrative includes at a minimum: • character **and** situation (activity **or** setting) • a conclusion that **may not** follow from the narrated experiences **or** events	The narrative includes at a minimum: • **some** evidence related to a character, situation (activity **or** setting), **or** conclusion	**0** • **no** evidence of organization	**5** • evidence is **off topic**
Idea Development – The narrative includes a description of events using concrete words or sensory details (e.g., how things look, sound, taste, smell, or feel) related to the events.	The narrative includes at a minimum: • **two** events related to the situation (activity **or** setting) • **both** of the events include a detail related to character's action **or** response to a situation (activity **or** setting)	The narrative includes at a minimum: • **two** events related to the situation (activity **or** setting) • **one** of the events includes a detail related to a character's action **or** response to a situation (activity **or** setting)	The narrative includes at a minimum: • **one** event related to the situation (activity **or** setting)	**0** • **no** evidence of idea development	**5** • evidence is **off topic**
Conventions – Students use standard English conventions (e.g., end punctuation, subject-verb agreement).	The essay includes **more than one sentence and** at a minimum: • end punctuation for **more than one** thought unit • **one** complex thought unit that expresses a complete idea **with** subject-verb agreement (e.g., "The dog runs" or "the dog runs")	The narrative includes at a minimum: • end punctuation for **one** thought unit • **one** complex thought unit **with or without** subject-verb agreement	The narrative includes at a minimum: • **one** use of standard English conventions (end punctuation for **one** thought unit **or one** thought unit **with or without** subject-verb agreement)	**0** • **no** evidence of standard English conventions	

Grade 4 Writing Scoring Rubric

Level 3

Rubric Elements	Full Evidence 3	Partial Evidence 2	Limited Evidence 1	Unrelated Evidence 0 or 5	
Organization – The narrative establishes a situation (activity and setting) and includes a character. The response provides a conclusion.	The narrative includes at a minimum: • character **and** situation (activity **and** setting) • description of character **and** situation (activity **or** setting) • a conclusion that follows from the narrated experiences **or** events	The narrative includes at a minimum: • character **and** situation (activity **or** setting) • description of the character **or** the situation (activity **or** setting) • a conclusion that **may not** follow from the narrated experiences **or** events	The narrative includes at a minimum: • **some** evidence related to a character, situation (activity **or** setting), **or** conclusion **OR** • descriptive words related to a character **or** situation (activity **or** setting)	**0** • **no** evidence of organization	**5** • evidence is **off topic**
Idea Development – The narrative includes a description of events using concrete words or sensory details (e.g., how things look, sound, taste, smell, or feel) related to the events.	The narrative includes at a minimum: • **two** events related to the situation (activity **or** setting) • **both** events include a detail related to a character's action **or** response to a situation (activity **or** setting)	The narrative includes at a minimum: • **two** events related to the situation (activity **or** setting) • **one** of the events includes a detail related to a character's action **or** response to a situation (activity **or** setting)	The narrative includes at a minimum: • **one** event related to the situation (activity **or** setting)	**0** • **no** evidence of idea development	**5** • evidence is **off topic**
Conventions – Students use standard English conventions (e.g., capitalization, end punctuation, subject-verb agreement).	The narrative includes at a minimum: • capitalization at the beginning of the **majority** of thought units • end punctuation for **more than one** thought unit • **one** complex thought unit that expresses a complete idea **with** subject-verb agreement (e.g., "The dog runs" or "the dog runs")	The narrative includes at a minimum: • capitalization at the beginning of **one** thought unit • end punctuation for **one** thought unit • **one** complex thought unit **with or without** subject-verb agreement	The narrative includes at a minimum: • **one** use of standard English conventions (capitalization at the beginning of **one** thought unit, end punctuation for **one** thought unit **or one** thought unit **with or without** subject-verb agreement)	**0** • **no** evidence of standard English conventions	

Grade 5 Writing Scoring Rubric

Level 2

Rubric Elements	Full Evidence 3	Partial Evidence 2	Limited Evidence 1	Unrelated Evidence 0 or 5	
Organization – The narrative establishes a situation (activity and setting) for the story and includes characters. The response provides a conclusion.	The narrative includes at a minimum: • **two** characters **unchanged** through the narrative • establish a situation (activity **and** setting) • a conclusion that follows from the narrated experiences **or** events	The narrative includes at a minimum: • **two** characters • a situation (activity **or** setting) • a conclusion that **may not** follow from the narrated experiences **or** events	The narrative includes at a minimum: • **some** evidence related to a character, situation (activity **or** setting), **or** conclusion	**0** • **no** evidence of organization	**5** • evidence is **off topic**
Idea Development – The narrative includes dialogue, and events supported with relevant details and descriptive statements.	The narrative includes at a minimum: • **two** events that connect to the narrative • **both** of the events include a detail related to a character's action **or** response to a situation (activity **or** setting) • **one** dialogue statement from one character to the other character **relevant to the narrative** (e.g., I said "No, I want to play.")	The narrative includes at a minimum: • **two** events related to a character's action **or** response to a situation (activity **or** setting) • **one** of the events includes a detail related to a character's action **or** response to a situation (activity **or** setting) • **one** dialogue statement from one character to the other character that **may not** be relevant to the narrative	The narrative includes at a minimum: • **one** event related to the situation (activity **or** setting)	**0** • **no** evidence of idea development	**5** • evidence is **off topic**
Conventions – Students use standard English conventions (e.g., end punctuation, subject-verb agreement).	The essay includes **more than one** sentence and at a minimum: • end punctuation for **more than one** thought unit • **one** complete sentence that expresses an idea **with** subject-verb agreement (e.g., "The dog runs**.**")	The narrative includes at a minimum: • end punctuation for **one** thought unit • **one** complete sentence **with or without** subject-verb agreement	The narrative includes at a minimum: • **one** use of standard English conventions (end punctuation for **one** thought unit **or one** thought unit **with or without** subject-verb agreement)	**0** • **no** evidence of standard English conventions	

Grade 5 Writing Scoring Rubric

Level 3

Rubric Elements	Full Evidence 3	Partial Evidence 2	Limited Evidence 1	Unrelated Evidence 0 or 5	
Organization – The narrative establishes a situation (activity and setting) for the story and includes characters. The response provides a conclusion.	The narrative includes at a minimum: • **two** characters unchanged through narrative • identification of the situation (activity **and** setting) • a conclusion that follows from the narrated experiences **or** events	The narrative includes at a minimum: • **two** characters • identification of the setting **or** the activity • a conclusion that **may not** follow from the narrated experiences **or** events	The narrative includes at a minimum: • **some** evidence related to a character **or** conclusion	**0** • **no** evidence of organization	**5** • evidence is **off topic**
Idea Development – The narrative includes dialogue, and events supported with relevant details and descriptive statements.	The narrative includes at a minimum: • **two** sequenced events related to the situation (activity **or** setting) • **both** events include a detail related to a character's action **or** response to a situation (activity **or** setting) • **one** relevant conversation between two characters (e.g., I said "No! I don't want to go to bed." Mom said "OK.")	The narrative includes at a minimum: • **two** events related to a character's action **or** response to a situation (activity **or** setting) • **one** event that includes a detail related to a character's action **or** response to a situation (activity **or** setting) • **one** relevant piece of dialogue showing what one character said to the other	The narrative includes at a minimum: • **one** event related to the situation (activity **or** setting)	**0** • **no** evidence of idea development	**5** • evidence is **off topic**
Conventions – Students use standard English conventions (e.g., capitalization, end punctuation, subject-verb agreement).	The narrative includes **more than one sentence and** at a minimum: • capitalization at the beginning of the **majority** of thought units • end punctuation for the **majority** of thought units • **one** complete sentence that expresses an idea **with** subject-verb agreement (e.g., "**T**he dog runs**.**")	The narrative includes at a minimum: • capitalization at the beginning of **one** thought unit • end punctuation for **one** thought unit • **one** complete sentence **with** subject-verb agreement	The narrative includes at a minimum: • **one** use of standard English conventions (capitalization at the beginning of **one** thought unit, end punctuation for **one** thought unit **or one** thought unit **with or without** subject-verb agreement)	**0** • **no** evidence of standard English conventions	

Grade 6 Writing Scoring Rubric

Level 2

Rubric Elements	Full Evidence 3	Partial Evidence 2	Limited Evidence 1	Unrelated Evidence 0 or 5	
Organization – The essay addresses a specified topic and is organized to describe two opposing conditions (e.g., compare/contrast).	The essay includes at a minimum: • an introduction that states the essay is about **two opposing conditions** • a body that includes: o **one** activity for **each** of the two opposing conditions; **and** o **one** activity common to **both** conditions • a conclusion that states **two opposing conditions or** summarizes the content	The essay includes at a minimum: • an introduction that states **one** activity **or** topic • a body that relates **two** conditions with activities • a conclusion that states **one** activity **or** the topic	The essay includes at a minimum: • **some** evidence related to the specified topic (i.e., introduction, compare/contrast relationship, **or** conclusion)	**0** • **no** evidence of organization	**5** • evidence is **off topic**
Idea Development – The essay develops a topic, and includes relevant facts and details to promote meaning and create clarity.	The essay includes at a minimum: • **three** activities, each with relevant details (the same detail may be used for all activities **if relevant to each**)	The essay includes at a minimum: • **one** activity with a relevant detail	The essay includes at a minimum: • **one** detail that describes an activity	**0** • **no** evidence of idea development	**5** • evidence is **off topic**
Conventions – Students use standard English conventions (e.g., end punctuation, subject-verb agreement).	The essay includes **more than one sentence and** at a minimum: • end punctuation for **more than one** thought unit • **one** complete sentence that expresses an idea **with** subject-verb agreement (e.g., "The dog runs**.**")	The essay includes at a minimum: • end punctuation for **one** thought unit • **one** complete sentence **with or without** subject-verb agreement	The essay includes at a minimum: • **one** use of standard English conventions (end punctuation for **one** thought unit **or one** thought unit **with or without** subject-verb agreement)	**0** • **no** evidence of standard English conventions	

Grade 6 Writing Scoring Rubric

Level 3

Rubric Elements	Full Evidence 3	Partial Evidence 2	Limited Evidence 1	Unrelated Evidence 0 or 5	
<u>Organization</u> – The essay addresses a specified topic and is organized to describe two opposing conditions (e.g., compare/contrast). The response provides a conclusion.	The essay includes at a minimum: • an introduction that presents the **<u>two</u>** opposing conditions • a body that includes: o **<u>one</u>** activity **<u>common to</u> <u>both</u>** conditions o **<u>one</u>** activity related to **<u>each</u>** of the two opposing conditions • a conclusion that states the **<u>two</u>** opposing conditions	The essay includes at a minimum: • an introduction that presents the topic • a body that includes: o **<u>one</u>** activity **<u>common to</u> <u>both</u>** conditions o **<u>one</u>** activity related to **<u>one of the two</u>** opposing conditions • a conclusion that states the topic	The essay includes at a minimum: • **<u>some</u>** evidence related to the specified topic (i.e., introduction, compare/contrast relationship, **<u>or</u>** conclusion)	**0** • **<u>no</u>** evidence of organization	**5** • evidence is **<u>off topic</u>**
<u>Idea Development</u> – The essay develops a topic, and includes relevant facts and details to promote meaning and create clarity.	The essay includes at a minimum: • **<u>one</u>** activity related to **<u>both</u>** conditions **<u>with</u>** a relevant detail • **<u>one</u>** activity related to **<u>each</u>** of the two opposing conditions, **<u>each with</u>** relevant details	The essay includes at a minimum: • **<u>two</u>** activities **<u>each with</u>** a relevant detail	The essay includes at a minimum: • **<u>one</u>** activity **OR** • **<u>one</u>** detail that describes an activity	**0** • **<u>no</u>** evidence of idea development	**5** • evidence is **<u>off topic</u>**
<u>Conventions</u> – Students use standard English conventions (e.g., capitalization, end punctuation, subject-verb agreement).	The essay includes **<u>more than one sentence and</u>** at a minimum: • capitalization at the beginning of the **<u>majority</u>** of thought units • end punctuation for the **<u>majority</u>** of thought units • **<u>one</u>** complete sentence that expresses an idea **<u>with</u>** subject-verb agreement (e.g., "**<u>T</u>**he dog runs**<u>.</u>**")	The essay includes at a minimum: • capitalization at the beginning of **<u>one</u>** thought unit • end punctuation for **<u>one</u>** thought unit • **<u>one</u>** complete sentence **<u>with</u>** subject-verb agreement	The essay includes at a minimum: • **<u>one</u>** use of standard English conventions (capitalization at the beginning of **<u>one</u>** thought unit, end punctuation for **<u>one</u>** thought unit **<u>or one</u>** thought unit **<u>with or without</u>** subject-verb agreement)	**0** • **<u>no</u>** evidence of standard English conventions	

Grade 7 Writing Scoring Rubric

Level 2

<table>
<tr><th rowspan="2">Rubric Elements</th><th rowspan="2">Full Evidence
3</th><th rowspan="2">Partial Evidence
2</th><th rowspan="2">Limited Evidence
1</th><th colspan="2">Unrelated Evidence
0 or 5</th></tr>
<tr><th>0</th><th>5</th></tr>
<tr><td><u>Organization</u> – The essay addresses a specified topic and is organized with an effect related directly to a cause (e.g., cause/ effect).</td><td>The essay includes at a minimum:
• an introduction that states the topic/cause
• a body that relates the effect to the provided cause
• a conclusion that states the essay is about a cause <u>and</u> its effect</td><td>The essay includes at a minimum:
• an introduction that states the topic/cause
• a body that includes an effect that <u>may not</u> relate to the provided cause
• a conclusion that states a cause <u>or</u> the effect</td><td>The essay includes at a minimum:
• <u>some</u> evidence related to the specified topic (i.e., introduction, cause/effect relationship, <u>or</u> conclusion)</td><td>• <u>no</u> evidence of organization</td><td>• evidence is <u>off topic</u></td></tr>
<tr><td><u>Idea Development</u> – The essay develops a topic, and includes details to promote meaning and create clarity.</td><td>The essay includes at a minimum:
• <u>one</u> relevant detail to describe the effect</td><td>The essay includes at a minimum:
• <u>one</u> effect with <u>no</u> relevant detail</td><td>The essay includes at a minimum:
• <u>one</u> idea related to the topic</td><td>0
• <u>no</u> evidence of idea development</td><td>5
• evidence is <u>off topic</u></td></tr>
<tr><td><u>Conventions</u> – Students use standard English conventions (e.g., end punctuation, subject-verb agreement).</td><td>The essay includes <u>more than one sentence</u> and at a minimum:
• end punctuation for <u>more than one</u> thought unit
• <u>one</u> complete sentence that expresses an idea <u>with</u> subject-verb agreement (e.g., “The dog runs.”)</td><td>The essay includes at a minimum:
• end punctuation for <u>one</u> thought unit
• <u>one</u> complete sentence <u>with or without</u> subject-verb agreement</td><td>The essay includes at a minimum:
• <u>one</u> use of standard English conventions (end punctuation for <u>one</u> thought unit <u>or one</u> thought unit <u>with or without</u> subject-verb agreement)</td><td colspan="2">0
• <u>no</u> evidence of standard English conventions</td></tr>
</table>

Grade 7 Writing Scoring Rubric

Level 3

Rubric Elements	Full Evidence 3	Partial Evidence 2	Limited Evidence 1	Unrelated Evidence 0 or 5	
				0	5
<u>Organization</u> – The essay addresses a specified topic and is organized with an effect related directly to a cause (e.g., cause/effect).	The essay includes at a minimum: • an introduction that presents the cause **<u>and</u>** its effects • a body that includes **<u>two</u>** effects **<u>and</u>** refers them to the cause • a conclusion that states the essay is about a cause **<u>and</u>** its effects	The essay includes at a minimum: • an introduction that presents a topic • a body that includes **<u>one</u>** effect **<u>and</u>** refers it to the cause • a conclusion that states the topic	The essay includes at a minimum: • **<u>some</u>** evidence related to the specified topic (i.e., introduction, on-topic cause/effect relationship, or conclusion)	• **<u>no</u>** evidence of organization	• evidence is **<u>off topic</u>**
<u>Idea Development</u> – The essay develops a topic, and includes details and transitional words to promote meaning and create clarity.	The essay includes at a minimum: • **<u>two</u>** effects, **<u>each</u>** with a relevant detail • transitional words to connect the cause to **<u>each</u>** of the **<u>two</u>** effects	The essay includes at a minimum: • **<u>one</u>** effect **<u>with</u>** a relevant detail • transitional word to **<u>connect</u>** **<u>one</u>** cause/effect relationship	The essay includes at a minimum: • **<u>one</u>** detail that describes the cause **<u>or</u>** effect **OR** • **<u>one</u>** transition word	• **<u>no</u>** evidence of idea development	• evidence is **<u>off topic</u>**
<u>Conventions</u> – Students use standard English conventions (e.g., capitalization, end punctuation, subject-verb agreement).	The essay includes **<u>more than one sentence and</u>** at a minimum: • capitalization at the beginning of the **<u>majority</u>** of thought units • end punctuation for the **<u>majority</u>** of thought units • **<u>one</u>** complete sentence that expresses an idea **<u>with</u>** subject-verb agreement (e.g., "<u>T</u>he dog runs<u>.</u>")	The essay includes at a minimum: • capitalization at the beginning of **<u>one</u>** thought unit • end punctuation for **<u>one</u>** thought unit • **<u>one</u>** complete sentence **<u>with</u>** subject-verb agreement	The essay includes at a minimum: • **<u>one</u>** use of standard English conventions (capitalization at the beginning of **<u>one</u>** thought unit, end punctuation for **<u>one</u>** thought unit **<u>or one</u>** thought unit **<u>with or without</u>** subject-verb agreement)	0 • **<u>no</u>** evidence of standard English conventions	

Grade 8 Writing Scoring Rubric

Level 2

Rubric Elements	Full Evidence 3	Partial Evidence 2	Limited Evidence 1	Unrelated Evidence 0 or 5	
				0	**5**
Organization – The essay addresses the specified topic and is organized with a solution related directly to the problem (e.g., problem/solution).	The essay includes at a minimum: • an introduction that states **both** parts of the problem • a body that relates **how** the solution can be applied to the problem • a conclusion that states the problem **and** the solution	The essay includes at a minimum: • an introduction that states the problem • **one** solution that **may not** relate to the problem • a conclusion that states the problem **or** the solution	The essay includes at a minimum: • **some** evidence related to the specified topic (i.e., introduction, on-topic problem/solution relationship, **or** conclusion)	• **no** evidence of organization	• evidence is **off topic**
Idea Development – The essay develops a topic, and includes details to promote meaning and create clarity.	The essay includes at a minimum: • **one** relevant detail to describe the problem • **one** relevant detail to describe the solution	The essay includes at a minimum: • **one** relevant detail to describe the problem **or** the solution	The essay includes at a minimum: • **one** detail **or** word that describes the problem **or** the solution	• **no** evidence of idea development	• evidence is **off topic**
Conventions – Students use standard English conventions (end punctuation, subject-verb agreement).	The essay includes **more than one sentence and** at a minimum: • end punctuation for **more than one** thought unit • **one** complete sentence that expresses an idea **with** subject-verb agreement (e.g., “The dog runs**.**”)	The essay includes at a minimum: • end punctuation for **one** thought unit • **one** complete sentence **with or without** subject-verb agreement	The essay includes at a minimum: • **one** use of standard English conventions (end punctuation for **one** thought unit **or one** thought unit **with or without** subject-verb agreement)	**0** • **no** evidence of standard English conventions	

Grade 8 Writing Scoring Rubric

Level 3

Rubric Elements	Full Evidence 3	Partial Evidence 2	Limited Evidence 1	Unrelated Evidence 0 or 5	
Organization – The essay addresses the specified topic and is organized with a solution related directly to the problem (e.g., problem/solution).	The essay includes at a minimum: • an introduction that states **both** parts of the problem • a body that includes a solution **and** refers to the problem • a conclusion that states the problem **and** its solution	The essay includes at a minimum: • an introduction that states **one** part of the problem • a body that includes a **related** solution • a conclusion that states the problem **or** the solution	The essay includes at a minimum: • **some** evidence related to the specified topic (i.e., introduction, on-topic problem/solution relationship, **or** conclusion)	**0** • **no** evidence of organization	**5** • evidence is **off topic**
Idea Development – The essay develops a topic, and includes details and transitional words to promote meaning and create clarity.	The essay includes at a minimum: • **one** problem **with** a relevant detail • **one** solution **with** a relevant detail • **one** transitional word(s) that connects the problem to the solution	The essay includes at a minimum: • **one** problem **or** solution **with** a relevant detail • **one** transitional word that is in relation to the problem **or** the solution	The essay includes at a minimum: • **one** detail **or** word that describes the problem **or** the solution	**0** • **no** evidence of idea development	**5** • evidence is **off topic**
Conventions – Students use standard English conventions (e.g., capitalization, end punctuation, subject-verb agreement).	The essay includes **more than one sentence and** at a minimum: • capitalization at the beginning of the **majority** of thought units • end punctuation for the **majority** of thought units • **one** complete sentence that expresses an idea **with** subject-verb agreement (e.g., "The dog runs.")	The essay includes at a minimum: • capitalization at the beginning of **one** thought unit • end punctuation for **one** thought unit • **one** complete sentence **with** subject-verb agreement	The essay includes at a minimum: • **one** use of standard English conventions (capitalization at the beginning of **one** thought unit, end punctuation for **one** thought unit **or one** thought unit **with or without** subject-verb agreement)	**0** • **no** evidence of standard English conventions	

High School Writing Scoring Rubric

Level 2

Rubric Elements	Full Evidence 3	Partial Evidence 2	Limited Evidence 1	Unrelated Evidence 0 or 5	
Organization – The essay addresses a specified claim supported with organized complex ideas.	The essay includes at a minimum: • an introduction that states the claim **and** a rational reason • a conclusion that states the claim **and** the rational reason	The essay includes at a minimum: • an introduction that states the claim **or** a reason • a conclusion that states the claim **or** the reason	The essay includes at a minimum: • **some** evidence related to the specified claim/topic (i.e., introduction, claim/topic, **or** conclusion)	0 • **no** evidence of organization	5 • evidence is **off topic**
Idea Development – The defended claim includes relevant evidence, and uses words, phrases, and clauses to clarify the relationship among claim, reasons, and evidence	The essay includes at a minimum: • a body with **two** relevant facts **or** examples • words **or** phrases to connect the reason with **one** relevant fact **or** example	The essay includes at a minimum: • a body with **one** relevant fact **or** example • **one** word **or** phrase to connect the reason with **one** fact or example	The essay includes at a minimum: • **one** word related to the reason	0 • **no** evidence of idea development	5 • evidence is **off topic**
Conventions – Students use standard English conventions (e.g., end punctuation, subject-verb agreement).	The essay includes **more than one sentence** and at a minimum: • end punctuation for **more than one** thought unit • **one** complete sentence that expresses an idea **with** subject-verb agreement (e.g., "The dog runs**.**")	The essay includes at a minimum: • end punctuation for **one** thought unit • **one** complete sentence **with or without** subject-verb agreement	The essay includes at a minimum: • **one** use of standard English conventions (end punctuation for **one** thought unit **or one** thought unit **with or without** subject-verb agreement)	0 • **no** evidence of standard English conventions	

High School Writing Scoring Rubric

Level 3

Rubric Elements	Full Evidence 3	Partial Evidence 2	Limited Evidence 1	Unrelated Evidence 0 or 5	
Organization – The essay addresses a specified claim supported with organized complex ideas.	The essay includes at a minimum: • an introduction that states the claim **and** is supported by **two** rational reasons • a body that includes **two** reasons related to the claim • a conclusion that states the claim **and** is supported by **two** rational reasons	The essay includes at a minimum: • an introduction that states the claim • a body that includes **one** reason related to the claim • a conclusion that states the claim **with one** rational reason **or** relevant evidence	The essay includes at a minimum: • **some** evidence related to the specified claim/topic (i.e., introduction, claim/topic, or conclusion)	0 • **no** evidence of organization	5 • evidence is **off topic**
Idea Development – The defended claim includes relevant evidence, and uses words, phrases, and clauses to clarify the relationship among claim, reasons, and evidence.	The essay includes at a minimum: • **one** piece of **relevant** evidence that follows **each of the two** provided reasons • words or phrases that **connect each of the two** reasons **with relevant** evidence	The essay includes at a minimum: • a body with **one** reason **and one** piece of relevant evidence • a word **or** phrase that connects **one** reason **with one** piece of **relevant** evidence	The essay includes at a minimum: • **one** word related to the reason **or** a connecting word or phrase	0 • **no** evidence of idea development	5 • evidence is **off topic**
Conventions – Students use standard English conventions (e.g., capitalization, end punctuation, subject-verb agreement).	The essay includes **more than one sentence and** at a minimum: • capitalization at the beginning of the **majority** of thought units • end punctuation for the **majority** of thought units • **one** complete sentence that expresses an idea **with** subject-verb agreement (e.g., “**T**he dog runs**.**”)	The essay includes at a minimum: • capitalization at the beginning of **one** thought unit • end punctuation for **one** thought unit • **one** complete sentence **with** subject-verb agreement	The essay includes at a minimum: • **one** use of standard English conventions (capitalization at the beginning of **one** thought unit, end punctuation for **one** thought unit **or one** thought unit **with or without** subject-verb agreement)	0 • **no** evidence of standard English conventions	

Appendix B: Performance Level Descriptors

Performance Level Descriptors for ELA, Mathematics, and Science

MSAA developed PLDs for ELA and mathematics at grades 3–8 and high school through an iterative process involving multiple stakeholder groups. Content and Accessibility specialists also collaborated with MSAA Science Partners to develop PLDs for science in grades 5, 8, and high school. The MSAA partnership developed grade-level PLDs to summarize the knowledge, skills, and abilities (KSAs) prioritized for the MSAA that students need to attain at each level of achievement (Level 1–Level 4). Each performance level is understood to include the KSAs of the preceding performance levels.

The PLDs included in this appendix provide a detailed description for teachers, parents, and the public to see not only what grade-level content a student should know and be able to do in order to meet high expectations, but also the depth, breadth, and complexity of that content.

By using the PLDs, test results become multi-dimensional. Test results in the form of scale scores are one way educators, parents, and guardians find out where a student's performance is in relation to other students. The PLDs provide another dimension that completes the description of how a student interacts with the standards the test measures. Both the scale score and the PLDs provide information that helps teachers, schools, parents, and guardians build a path to student learning.

Grade 3 ELA Performance Level Descriptors

Level 1	Level 2*	Level 3*	Level 4*
Low text complexity – *Brief text with straightforward ideas and relationships; short, simple sentences*	**Low text complexity –** *Brief text with straightforward ideas and relationships; short, simple sentences*	**Moderate text complexity –** *Text with clear, complex ideas and relationships and simple, compound sentences*	**High text complexity –** *Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words*
In reading, the student is able to: • identify the topic of a literary text • identify a detail from a literary text • identify a character or setting in a literary text • identify the topic of an informational text • identify a title, caption, or heading in an informational text • identify an illustration related to a given topic • identify a topic presented by an illustration • identify the meaning of words (i.e., nouns)	**In reading, the student is able to:** • determine the central idea and supporting details in literary text • determine the main idea and identify supporting details in informational text • determine the main idea of visually presented information • identify the purpose of text features in informational text • use information from charts, graphs, diagrams, or timelines in informational text to answer questions • use context to identify the meaning of multiple-meaning words	**In reading, the student is able to:** • determine the central idea and supporting details in literary text • determine the main idea and identify supporting details in informational text • determine the main idea of visually presented information • identify the purpose of text features in informational text • use information from charts, graphs, diagrams, or timelines in informational text to answer questions • use context to identify the meaning of multiple-meaning words	**In reading, the student is able to:** • determine the central idea and supporting details in literary text • determine the main idea and identify supporting details in informational text • determine the main idea of visually presented information • identify the purpose of text features in informational text • use information from charts, graphs, diagrams, or timelines in informational text to answer questions • use context to identify the meaning of multiple-meaning words
	AND with Moderate text complexity – *Text with clear, complex ideas and relationships and simple, compound sentences*	**AND with High text complexity –** *Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words*	
	• use details from a literary text to answer specific questions • describe the relationship between characters, and character and setting in literary text	• use details from a literary text to answer specific questions • describe the relationship between characters, and character and setting in literary text	
	AND with accuracy, the student is able to: • identify simple words (i.e., words with a consonant at the beginning, a consonant at the end, and a short vowel in the middle)	**AND with accuracy, the student is able to:** • identify grade-level words	
AND in writing, the student is able to: • identify a statement related to an everyday topic • use the writing process to create a narrative product and demonstrate minimal (or no) command of organization, idea development, and/or conventions	**AND in writing, the student is able to:** • identify elements of a narrative text to include beginning, middle, and end • identify the category related to a set of facts • use the writing process to create a narrative product and demonstrate limited command of organization, idea development, and/or conventions	**AND in writing, the student is able to:** • identify a text feature (e.g., captions, graphs, or diagrams) to present information in explanatory text • use the writing process to create a narrative product and demonstrate partial command of organization, idea development, and/or conventions	**AND in writing, the student is able to:** • use the writing process to create a narrative product and demonstrate overall command of organization, idea development, and/or conventions

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

Grade 4 ELA Performance Level Descriptors

<table>
<tr><th>Level 1</th><th>Level 2*</th><th>Level 3*</th><th>Level 4*</th></tr>
<tr>
<td>Low text complexity –
Brief text with straightforward ideas and relationships; short, simple sentences</td>
<td>Low text complexity –
Brief text with straightforward ideas and relationships; short, simple sentences</td>
<td>Moderate text complexity –
Text with clear, complex ideas and relationships and simple, compound sentences</td>
<td>High text complexity –
Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words</td>
</tr>
<tr>
<td rowspan="4">In reading, the student is able to:
• identify a topic of a literary text
• identify a detail from a literary text
• identify a character in a literary text
• identify charts, graphs, diagrams, or timelines in an informational text
• identify a topic of an informational text
• use context to identify the meaning of multiple-meaning words
• identify general academic words</td>
<td>In reading, the student is able to:
• determine the theme of literary text and identify supporting details
• describe character traits using text-based details in literary text
• determine the main idea of informational text
• locate information in charts, graphs, diagrams, or timelines
• use information from charts, graphs, diagrams, or timelines in informational text to answer questions
• use general academic words</td>
<td>In reading, the student is able to:
• determine the theme of literary text and identify supporting details
• determine the main idea of informational text
• explain how the information provided in charts, graphs, diagrams, or timelines contributes to an understanding of informational text
• use information from charts, graphs, diagrams, or timelines in informational text to answer questions
• use general academic words</td>
<td rowspan="4">In reading, the student is able to:
• determine the theme of literary text and identify supporting details
• determine the main idea of informational text
• explain how the information provided in charts, graphs, diagrams, or timelines contributes to an understanding of informational text
• use information from charts, graphs, diagrams, or timelines in informational text to answer questions
• use general academic words</td>
</tr>
<tr>
<td>AND with Moderate text complexity –
Text with clear, complex ideas and relationships and simple, compound sentences</td>
<td>AND with High text complexity –
Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words</td>
</tr>
<tr>
<td>• use details from a literary text to answer specific questions
• use context to identify the meaning of multiple-meaning words</td>
<td>• use details from a literary text to answer specific questions
• describe character traits using text-based details in literary text
• use context to identify the meaning of multiple-meaning words</td>
</tr>
<tr>
<td>AND with accuracy, the student is able to:
• identify simple words (i.e., words with a consonant at the beginning, a consonant at the end, and a short vowel in the middle)</td>
<td>AND with accuracy, the student is able to:
• identify grade-level words</td>
</tr>
<tr>
<td>AND in writing, the student is able to:
• identify the concluding sentence in a short explanatory text
• use the writing process to create a narrative product and demonstrate minimal (or no) command of organization, idea development, and/or conventions</td>
<td>AND in writing, the student is able to:
• identify elements of a narrative text to include beginning, middle, and end
• identify a concluding sentence related to information in explanatory text
• use the writing process to create a narrative product and demonstrate limited command of organization, idea development, and/or conventions</td>
<td>AND in writing, the student is able to:
• identify a text feature (e.g., headings, charts, or diagrams) to present information in explanatory text
• use the writing process to create a narrative product and demonstrate partial command of organization, idea development, and/or conventions</td>
<td>AND in writing, the student is able to:
• use the writing process to create a narrative product and demonstrate overall command of organization, idea development, and/or conventions</td>
</tr>
</table>

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

Grade 5 ELA Performance Level Descriptors

Level 1	Level 2*	Level 3*	Level 4*
Low text complexity – *Brief text with straightforward ideas and relationships; short, simple sentences*	**Low text complexity –** *Brief text with straightforward ideas and relationships; short, simple sentences*	**Moderate text complexity –** *Text with clear, complex ideas and relationships and simple, compound sentences*	**High text complexity –** *Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words*
In reading, the student is able to: • identify an event from the beginning of a literary text • identify a detail from a literary text • identify a character, setting, and event in a literary text • identify the topic of an informational text • identify the main idea of an informational text • identify the difference in how information is presented in two sentences	**In reading, the student is able to:** • compare characters, settings, and events in literary text • determine the main idea and identify supporting details in informational text • use details from the text to support an author's point in informational text • compare and contrast how information and events are presented in two informational texts • use context to identify the meaning of multiple-meaning words	**In reading, the student is able to:** • compare characters, settings, and events in literary text • determine the main idea and identify supporting details in informational text • use details from the text to support an author's point in informational text • compare and contrast how information and events are presented in two informational texts • use context to identify the meaning of multiple-meaning words	**In reading, the student is able to:** • compare characters, settings, and events in literary text • determine the main idea and identify supporting details in informational text • use details from the text to support an author's point in informational text • compare and contrast how information and events are presented in two informational texts • use context to identify the meaning of multiple-meaning words
	AND with Moderate text complexity – *Text with clear, complex ideas and relationships and simple, compound sentences*	**AND with High text complexity –** *Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words*	
	• summarize a literary text from beginning to end • use details from a literary text to answer specific questions	• summarize a literary text from beginning to end • use details from a literary text to answer specific questions	
AND in writing, the student is able to: • identify the category related to a set of common nouns • use the writing process to create a narrative product and demonstrate minimal (or no) command of organization, idea development, and/or conventions	**AND in writing, the student is able to:** • identify elements of a narrative text to include beginning, middle, and end • identify a sentence that is organized for a text structure such as comparison/contrast • use the writing process to create a narrative product and demonstrate limited command of organization, idea development, and/or conventions	**AND in writing, the student is able to:** • support an explanatory text topic with relevant information • use the writing process to create a narrative product and demonstrate partial command of organization, idea development, and/or conventions	**AND in writing, the student is able to:** • use the writing process to create a narrative product and demonstrate overall command of organization, idea development, and/or conventions

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

Grade 6 ELA Performance Level Descriptors

<table>
<tr><th>Level 1</th><th>Level 2*</th><th>Level 3*</th><th>Level 4*</th></tr>
<tr>
<td>Low text complexity –
Brief text with straightforward ideas and relationships; short, simple sentences</td>
<td>Low text complexity –
Brief text with straightforward ideas and relationships; short, simple sentences</td>
<td>Moderate text complexity –
Text with clear, complex ideas and relationships and simple, compound sentences</td>
<td>High text complexity –
Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words</td>
</tr>
<tr>
<td rowspan="3">In reading, the student is able to:
• identify an event from the beginning or end of a literary text
• identify a detail from a literary text
• identify a character in a literary text
• identify the topic of an informational text
• identify the main idea of an informational text
• identify a fact from an informational text
• identify a description of an individual or event in an informational text
• use context to identify the meaning of multiple-meaning words
• identify the meaning of general academic words</td>
<td>In reading, the student is able to:
• summarize a literary text from beginning to end without including personal opinions
• support inferences about characters using details in literary text
• use details from the text to elaborate a key idea in informational text</td>
<td>In reading, the student is able to:
• summarize a literary text from beginning to end without including personal opinions
• support inferences about characters using details in literary text
• summarize an informational text without including personal opinions
• use details from the text to elaborate a key idea in informational text
• use evidence from the text to support an author’s claim in informational text
• summarize information presented in two informational texts
• use domain-specific words accurately</td>
<td rowspan="3">In reading, the student is able to:
• summarize a literary text from beginning to end without including personal opinions
• use details from a literary text to answer specific questions
• support inferences about characters using details in literary text
• use details from the text to elaborate a key idea in an informational text
• use evidence from the text to support an author’s claim in informational text
• use domain-specific words accurately</td>
</tr>
<tr>
<td>AND with Moderate text complexity –
Text with clear, complex ideas and relationships and simple, compound sentences</td>
<td>AND with High text complexity –
Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words</td>
</tr>
<tr>
<td>• use details from a literary text to answer specific questions
• use context to identify the meaning of multiple-meaning words</td>
<td>• use details from a literary text to answer specific questions
• use context to identify the meaning of multiple-meaning words</td>
</tr>
<tr>
<td>AND in writing, the student is able to:
• identify an everyday order of events
• use the writing process to create an explanatory product and demonstrate minimal (or no) command of organization, idea development, and/or conventions</td>
<td>AND in writing, the student is able to:
• identify elements of an explanatory text to include introduction, body, and conclusion
• identify the next event in a brief narrative
• use the writing process to create an explanatory product and demonstrate limited command of organization, idea development, and/or conventions</td>
<td>AND in writing, the student is able to:
• identify transition words and phrases to convey a sequence of events in narrative text
• use the writing process to create an explanatory product and demonstrate partial command of organization, idea development, and/or conventions</td>
<td>AND in writing, the student is able to:
• use the writing process to create an explanatory product and demonstrate overall command of organization, idea development, and/or conventions</td>
</tr>
</table>

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

Grade 7 ELA Performance Level Descriptors

<table>
<tr><th>Level 1</th><th>Level 2*</th><th>Level 3*</th><th>Level 4*</th></tr>
<tr>
<td>Low text complexity –
Brief text with straightforward ideas and relationships; short, simple sentences</td>
<td>Low text complexity –
Brief text with straightforward ideas and relationships; short, simple sentences</td>
<td>Moderate text complexity –
Text with clear, complex ideas and relationships and simple, compound sentences</td>
<td>High text complexity –
Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words</td>
</tr>
<tr>
<td rowspan="3">In reading, the student is able to:
• identify a theme from a literary text
• identify an inference from a literary text
• identify a conclusion from an informational text
• identify a claim the author makes in an informational text
• compare and contrast two statements related to the same topic
• use context to identify the meaning of words</td>
<td>In reading, the student is able to:
• identify the relationship between individuals or events in an informational text
• use evidence from the text to support an author’s claim in informational text</td>
<td>In reading, the student is able to:
• use details to support a conclusion from informational text
• use details to explain how the interactions between individuals, events, or ideas in informational texts are influenced by each other
• use evidence from the text to support an author’s claim in informational text
• compare and contrast how two authors write about the same topic in informational texts
• use context to identify the meaning of grade-level phrases</td>
<td rowspan="3">In reading, the student is able to:
• use details to support a conclusion from informational text
• use details to explain how the interactions between individuals, events, or ideas in informational texts are influenced by each other
• use evidence from the text to support an author’s claim in informational text
• compare and contrast how two authors write about the same topic in informational texts
• use context to identify the meaning of grade-level phrases</td>
</tr>
<tr>
<td>AND with Moderate text complexity –
Text with clear, complex ideas and relationships and simple, compound sentences</td>
<td>AND with High text complexity –
Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words</td>
</tr>
<tr>
<td>• use details to support themes from literary text
• use details to support inferences from literary text</td>
<td>• use details to support themes from literary text
• use details to support inferences from literary text</td>
</tr>
<tr>
<td>AND in writing, the student is able to:
• identify a graphic that includes an event as described in a text
• use the writing process to create an explanatory product and demonstrate minimal (or no) command of organization, idea development, and/or conventions</td>
<td>AND in writing, the student is able to:
• identify elements of an explanatory text to include introduction, body, and conclusion
• identify the next event in a brief narrative
• use the writing process to create an explanatory product and demonstrate limited command of organization, idea development, and/or conventions</td>
<td>AND in writing, the student is able to:
• identify a sentence that provides a conclusion in narrative text
• use the writing process to create an explanatory product and demonstrate partial command of organization, idea development, and/or conventions</td>
<td>AND in writing, the student is able to:
• use the writing process to create an explanatory product and demonstrate overall command of organization, idea development, and/or conventions</td>
</tr>
</table>

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

Grade 8 ELA Performance Level Descriptors

Level 1	Level 2*	Level 3*	Level 4*
Low text complexity – *Brief text with straightforward ideas and relationships; short, simple sentences*	**Low text complexity –** *Brief text with straightforward ideas and relationships; short, simple sentences*	**Moderate text complexity –** *Text with clear, complex ideas and relationships and simple, compound sentences*	**High text complexity –** *Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words*
In reading, the student is able to: • identify a theme from a literary text • identify an inference from a literary text • identify a fact related to a presented argument in informational text • identify a similar topic in two informational texts • use context to identify the meaning of multiple-meaning words • identify the meaning of general academic words	**In reading, the student is able to:** • use details to support a conclusion from literary text • identify an inference drawn from an informational text • identify the portion of text that contains specific information • identify an argument the author makes in informational text • examine parts of two informational texts to identify where the texts disagree on matters of fact or interpretation • use domain-specific words or phrases accurately	**In reading, the student is able to:** • use details to support a conclusion from literary text • use details to support an inference from informational text • identify the information (e.g., facts or quotes) in a section of text that contributes to the development of an idea • identify an argument the author makes in informational text • examine parts of two informational texts to identify where the texts disagree on matters of fact or interpretation • use domain-specific words and phrases accurately	**In reading, the student is able to:** • use details to support a conclusion from literary text • use details to support an inference from informational text • identify the information (e.g., facts or quotes) in a section of text that contributes to the development of an idea • identify an argument the author makes in informational text • examine parts of two informational texts to identify where the texts disagree on matters of fact or interpretation • use domain-specific words and phrases accurately
	AND with Moderate text complexity – *Text with clear, complex ideas and relationships and simple, compound sentences*	**AND with High text complexity –** *Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words*	
	• analyze the development of a theme including the relationship between a character and an event in literary text • use context to identify the meaning of grade-level words and phrases	• analyze the development of a theme including the relationship between a character and an event in literary text • use context to identify the meaning of grade-level words and phrases	
AND in writing, the student is able to: • identify a writer's opinion • use the writing process to create an explanatory product and demonstrate minimal (or no) command of organization, idea development, and/or conventions	**AND in writing, the student is able to:** • identify elements of an explanatory text to include introduction, body, and conclusion • identify an idea relevant to a claim • use the writing process to create an explanatory product and demonstrate limited command of organization, idea development, and/or conventions	**AND in writing, the student is able to:** • identify relevant information to support a claim • use the writing process to create an explanatory product and demonstrate partial command of organization, idea development, and/or conventions	**AND in writing, the student is able to:** • use the writing process to create an explanatory product and demonstrate overall command of organization, idea development, and/or conventions

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

High School ELA Performance Level Descriptors

<table>
<tr><th>Level 1</th><th>Level 2*</th><th>Level 3*</th><th>Level 4*</th></tr>
<tr>
<td>Low text complexity –
Brief text with straightforward ideas and relationships; short, simple sentences</td>
<td>Low text complexity –
Brief text with straightforward ideas and relationships; short, simple sentences</td>
<td>Moderate text complexity –
Text with clear, complex ideas and relationships and simple, compound sentences</td>
<td>High text complexity –
Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words</td>
</tr>
<tr>
<td rowspan="3">In reading, the student is able to:
• identify a summary of a literary text
• identify an event from a literary text
• identify the central idea of an informational text
• identify facts from an informational text
• identify what an author tells about a topic in informational text
• use context to identify the meaning of multiple-meaning words
• identify a word used to describe a person, place, thing, action, or event</td>
<td>In reading, the student is able to:
• use details to support a summary of literary text
• identify a conclusion from an informational text
• identify key details that support the development of a central idea of an informational text
• use details presented in two informational texts to answer a question
• explain why an author uses specific word choices within texts</td>
<td>In reading, the student is able to:
• use details to support a summary of literary text
• use details to support a conclusion presented in informational text
• identify key details that support the development of a central idea of an informational text
• use details presented in two informational texts to answer a question
• explain why an author uses specific word choices within texts</td>
<td rowspan="3">In reading, the student is able to:
• use details to support a summary of literary text
• use details to support a conclusion presented in informational text
• identify key details that support the development of a central idea of an informational text
• use details presented in two informational texts to answer a question
• explain why an author uses specific word choices within texts</td>
</tr>
<tr>
<td>AND with Moderate text complexity –
Text with clear, complex ideas and relationships and simple, compound sentences</td>
<td>AND with High text complexity –
Text with detailed and implied complex ideas and relationships; a variety of sentence types including phrases and transition words</td>
</tr>
<tr>
<td>• evaluate how the author’s use of specific details in literary text contributes to the text
• determine an author’s point of view about a topic in informational text
• use context to identify the meaning of grade-level phrases</td>
<td>• evaluate how the author’s use of specific details in literary text contributes to the text
• determine an author’s point of view about a topic in informational text
• use context to identify the meaning of grade-level phrases</td>
</tr>
<tr>
<td>AND in writing, the student is able to:
• identify information that is unrelated to a given topic
• use the writing process to create an argumentative product and demonstrate minimal (or no) command of organization, idea development, and/or conventions</td>
<td>AND in writing, the student is able to:
• identify elements of an argument to include introduction, claim, evidence, and conclusion
• identify how to group information for a specific text structure
• use the writing process to create an argumentative product and demonstrate limited command of organization, idea development, and/or conventions</td>
<td>AND in writing, the student is able to:
• identify relevant information to address a given topic and support the purpose of a text
• use the writing process to create an argumentative product and demonstrate partial command of organization, idea development, and/or conventions</td>
<td>AND in writing, the student is able to:
• use the writing process to create an argumentative product and demonstrate overall command of organization, idea development, and/or conventions</td>
</tr>
</table>

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

[Page is intentionally blank]

Grade 3 Mathematics Performance Level Descriptors

Level 1	Level 2*	Level 3*	Level 4*
Low task complexity – *Simple problems using common mathematical terms and symbols*	**Low task complexity –** *Simple problems using common mathematical terms and symbols*	**Moderate task complexity –** *Common problems presented in mathematical context using various mathematical terms and symbols*	**High task complexity –** *Multiple mathematical ideas presented in problems using various mathematical terms and symbolic representations of numbers, variables, and other item elements*
The student is able to: • solve addition problems • identify growing number patterns • identify an object showing a specified number of parts shaded • identify which object has the greater number of parts shaded • identify an object equally divided into two parts • identify the number of objects to be represented in a pictograph	**The student is able to:** • solve addition and subtraction word problems • identify an arrangement of objects that represents factors in a problem • solve multiplication equations in which both numbers are equal to or less than 5 • identify multiplication patterns • identify a set of objects as nearer to 1 or 10 • identify a representation of the area of a rectangle	**The student is able to:** • solve addition and subtraction word problems • check the correctness of an answer in the context of a scenario • solve multiplication equations in which both numbers are equal to or less than 5 • identify multiplication patterns • match fraction models to unitary fractions • compare fractions with different numerators and the same denominator • transfer data from an organized list to a bar graph	**The student is able to:** • solve addition and subtraction word problems • check the correctness of an answer in the context of a scenario • solve multiplication equations in which both numbers are equal to or less than 5 • identify multiplication patterns • match fraction models to unitary fractions • compare fractions with different numerators and the same denominator • transfer data from an organized list to a bar graph
	AND with Moderate task complexity – *Common problems presented in mathematical context using various mathematical terms and symbols*	**AND with High task complexity –** *Common problems presented in mathematical context using various mathematical terms and symbols*	
	• identify geometric figures that are divided into equal parts	• round numbers to the nearest 10 • identify geometric figures that are divided into equal parts • count unit squares to compute the area of a rectangle	

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

Grade 4 Mathematics Performance Level Descriptors

Level 1	Level 2*	Level 3*	Level 4*
Low task complexity – *Simple problems using common mathematical terms and symbols*	**Low task complexity –** *Simple problems using common mathematical terms and symbols*	**Moderate task complexity –** *Common problems presented in mathematical context using various mathematical terms and symbols*	**High task complexity –** *Multiple mathematical ideas presented in problems using various mathematical terms and symbolic representations of numbers, variables, and other item elements*
The student is able to: • identify an array with the same number of objects in each row • identify values rounded to the nearest tens place • identify equivalent representations of a fraction (e.g., shaded diagram) • compare representations of a fraction (e.g., shaded diagram) • identify a rectangle with the larger or smaller perimeter • identify a given attribute of a shape • identify the data drawn in a bar graph that represents the greatest value	**The student is able to:** • match a model to a multiplication expression using two single-digit numbers • identify a model of a multiplicative comparison • show division of objects into equal groups • round numbers to the nearest 10, 100, or 1000 • differentiate parts and wholes • compute the perimeter of a rectangle	**The student is able to:** • solve multiplication word problems • show division of objects into equal groups • round numbers to the nearest 10, 100, or 1000 • compare two fractions with different denominators • sort a set of two-dimensional shapes • compute the perimeter of a rectangle • transfer data to a graph	**The student is able to:** • solve multiplication word problems • show division of objects into equal groups • round numbers to the nearest 10, 100, or 1000 • compare two fractions with different denominators • sort a set of two-dimensional shapes • compute the perimeter of a rectangle • transfer data to a graph
	AND with Moderate task complexity – *Common problems presented in mathematical context using various mathematical terms and symbols*	**AND with High task complexity –** *Common problems presented in mathematical context using various mathematical terms and symbols*	
	• identify equivalent fractions • select a two-dimensional shape with a given attribute	• solve a multiplicative comparison word problem using up to two-digit numbers • check the correctness of an answer in the context of a scenario • identify equivalent fractions	

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

Grade 5 Mathematics Performance Level Descriptors

<table>
<tr><th>Level 1</th><th>Level 2*</th><th>Level 3*</th><th>Level 4*</th></tr>
<tr>
<td>Low task complexity –
Simple problems using common mathematical terms and symbols</td>
<td>Low task complexity –
Simple problems using common mathematical terms and symbols</td>
<td>Moderate task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
<td>High task complexity –
Multiple mathematical ideas presented in problems using various mathematical terms and symbolic representations of numbers, variables, and other item elements</td>
</tr>
<tr>
<td rowspan="3">The student is able to:
solve one-step subtraction word problems
divide sets (no greater than 6) into two equal parts
identify values in the tenths place
identify a number in the ones, tens, or hundreds place
identify a given axis of a coordinate plane
match the conversion of 3 feet to 1 yard to a model
calculate elapsed time (i.e., hours)
identify whether the values increase or decrease in a line graph</td>
<td>The student is able to:
identify if the total will increase or decrease when combining sets
perform operations with decimals
identify a symbolic representation of the addition of two fractions
identify place values to the hundredths place
convert standard measurements</td>
<td>The student is able to:
solve multiplication and division word problems
perform operations with decimals
solve word problems involving fractions
identify place values to the hundredths place
locate a given point on a coordinate plane when given an ordered pair
convert standard measurements
convert between minutes and hours
make quantitative comparisons between data sets shown as line graphs</td>
<td rowspan="3">The student is able to:
solve multiplication and division word problems
perform operations with decimals
solve word problems involving fractions
identify place values to the hundredths place
locate a given point on a coordinate plane when given an ordered pair
convert standard measurements
convert between minutes and hours
make quantitative comparisons between data sets shown as line graphs</td>
</tr>
<tr>
<td>AND with Moderate task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
<td>AND with High task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
</tr>
<tr>
<td>compare the values of two products based upon multipliers
round decimals to the nearest whole number</td>
<td>compare the values of two products based upon multipliers
round decimals to the nearest whole number</td>
</tr>
</table>

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

Grade 6 Mathematics Performance Level Descriptors

<table>
<tr><th>Level 1</th><th>Level 2*</th><th>Level 3*</th><th>Level 4*</th></tr>
<tr>
<td>Low task complexity –
Simple problems using common mathematical terms and symbols</td>
<td>Low task complexity –
Simple problems using common mathematical terms and symbols</td>
<td>Moderate task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
<td>High task complexity –
Multiple mathematical ideas presented in problems using various mathematical terms and symbolic representations of numbers, variables, and other item elements</td>
</tr>
<tr>
<td rowspan="3">The student is able to:
• identify a model of a given percent
• match a given unit rate to a model
• identify a representation of two equal sets
• identify a number less than 0 on a number line
• identify the meaning of an unknown in a modeled equation
• count the number of grids or tiles inside a rectangle to find the area of a rectangle
• identify the object that appears most frequently in a set of data (mode)
• identify a representation of a set of data arranged into even groups (mean)</td>
<td>The student is able to:
• match a given ratio to a model
• recognize a representation of the sum of two halves
• solve real-world measurement problems involving unit rates
• identify a representation of a value less than 0
• identify the median or the equation needed to determine the mean of a set of data</td>
<td>The student is able to:
• perform operations using up to three-digit numbers
• solve real-world measurement problems involving unit rates
• identify positive and negative values on a number line
• determine the meaning of a value from a set of positive and negative integers
• solve word problems with expressions including variables
• compute the area of a parallelogram
• identify the median or the equation needed to determine the mean of a set of data</td>
<td rowspan="3">The student is able to:
• solve real-world measurement problems involving unit rates
• identify positive and negative values on a number line
• solve word problems with expressions including variables
• compute the area of a parallelogram
• identify the median or the equation needed to determine the mean of a set of data</td>
</tr>
<tr>
<td>AND with Moderate task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
<td>AND with High task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
</tr>
<tr>
<td>• perform one-step operations with two decimal numbers
• solve word problems using a percent</td>
<td>• perform one-step operations with two decimal numbers
• solve word problems using a percent
• solve word problems using ratios and rates</td>
</tr>
</table>

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

Grade 7 Mathematics Performance Level Descriptors

<table>
<tr><th>Level 1</th><th>Level 2*</th><th>Level 3*</th><th>Level 4*</th></tr>
<tr>
<td>Low task complexity –
Simple problems using common mathematical terms and symbols</td>
<td>Low task complexity –
Simple problems using common mathematical terms and symbols</td>
<td>Moderate task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
<td>High task complexity –
Multiple mathematical ideas presented in problems using various mathematical terms and symbolic representations of numbers, variables, and other item elements</td>
</tr>
<tr>
<td rowspan="3">The student is able to:
• identify a representation that represents a negative number and its multiplication or division by a positive number
• identify representations of area and circumference of a circle
• identify representations of surface area
• make qualitative comparisons when interpreting a data set presented on a bar graph or in a table</td>
<td>The student is able to:
• match a given ratio to a model
• identify the meaning of an unknown in a modeled equation
• describe a directly proportional relationship (i.e., increases or decreases)
• find the surface area of a three-dimensional right prism</td>
<td>The student is able to:
• solve division problems with positive/negative whole numbers
• solve word problems involving ratios
• use a proportional relationship to solve a percentage problem
• identify proportional relationships between quantities represented in a table
• identify unit rate (constant of proportionality) in tables and graphs of proportional relationships
• compute the area of a circle
• find the surface area of a three-dimensional right prism</td>
<td rowspan="3">The student is able to:
• solve division problems with positive/negative whole numbers
• solve word problems involving ratios
• identify proportional relationships between quantities represented in a table
• compute the area of a circle
• find the surface area of a three-dimensional right prism</td>
</tr>
<tr>
<td>AND with Moderate task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
<td>AND with High task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
</tr>
<tr>
<td>• solve multiplication problems with positive/negative whole numbers
• interpret graphs to qualitatively contrast data sets</td>
<td>• solve multiplication problems with positive/negative whole numbers
• evaluate variable expressions that represent word problems
• interpret graphs to qualitatively contrast data sets</td>
</tr>
</table>

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

Grade 8 Mathematics Performance Level Descriptors

<table>
<tr><th>Level 1</th><th>Level 2*</th><th>Level 3*</th><th>Level 4*</th></tr>
<tr>
<td>Low task complexity –
Simple problems using common mathematical terms and symbols</td>
<td>Low task complexity –
Simple problems using common mathematical terms and symbols</td>
<td>Moderate task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
<td>High task complexity –
Multiple mathematical ideas presented in problems using various mathematical terms and symbolic representations of numbers, variables, and other item elements</td>
</tr>
<tr>
<td rowspan="3">The student is able to:
• locate a given decimal number on a number line
• identify the relatively larger data set when given two data sets presented in a graph
• identify congruent rectangles
• identify similar rectangles
• identify an attribute of a cylinder
• identify a rectangle with the larger or smaller area as compared to another rectangle
• identify an ordered pair and its point on a graph</td>
<td>The student is able to:
• identify the solution to an equation that contains a variable
• identify the y-intercept of a linear graph
• match a given relationship between two variables to a model
• identify a data display that represents a given situation
• interpret data presented in graphs to identify associations between variables</td>
<td>The student is able to:
• locate approximate placement of an irrational number on a number line
• solve a linear equation that contains a variable
• identify the relationship shown on a linear graph
• calculate slope of a positive linear graph
• compute the change in area of a figure when its dimensions are changed
• solve for the volume of a cylinder
• plot provided data on a graph</td>
<td rowspan="3">The student is able to:
• locate approximate placement of an irrational number on a number line
• solve a linear equation that contains a variable
• identify the relationship shown on a linear graph
• compute the change in area of a figure when its dimensions are changed
• plot provided data on a graph</td>
</tr>
<tr>
<td>AND with Moderate task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
<td>AND with High task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
</tr>
<tr>
<td>• identify congruent figures
• use properties of similarity to identify similar figures
• interpret data tables to identify the relationship between variables</td>
<td>• interpret data presented in graphs to identify associations between variables
• interpret data tables to identify the relationship between variables
• use properties of similarity to identify similar figures
• identify congruent figures</td>
</tr>
</table>

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

High School Mathematics Performance Level Descriptors

<table>
<tr><th>Level 1</th><th>Level 2*</th><th>Level 3*</th><th>Level 4*</th></tr>
<tr>
<td>Low task complexity –
Simple problems using common mathematical terms and symbols</td>
<td>Low task complexity –
Simple problems using common mathematical terms and symbols</td>
<td>Moderate task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
<td>High task complexity –
Multiple mathematical ideas presented in problems using various mathematical terms and symbolic representations of numbers, variables, and other item elements</td>
</tr>
<tr>
<td rowspan="3">The student is able to:
arrange a given number of objects into two sets in multiple combinations
match an equation with a variable to a provided real-world situation
determine whether a given point is or is not part of a data set shown on a graph
identify an extension of a linear graph
use a table to match a unit conversion
complete the formula for area of a figure</td>
<td>The student is able to:
identify the model that represents a square number
identify variable expressions that represent word problems
identify the hypotenuse of a right triangle
identify the greatest or least value in a set of data shown on a number line
identify the missing label on a histogram
calculate the mean and median of a set of data</td>
<td>The student is able to:
compute the value of an expression that includes an exponent
identify variable expressions that represent word problems
solve real-world measurement problems that require unit conversions
find the missing attribute of a three-dimensional figure
determine two similar right triangles when a scale factor is given
make predictions from data tables and graphs to solve problems
plot data on a histogram
calculate the mean and median of a set of data</td>
<td rowspan="3">The student is able to:
identify variable expressions that represent word problems
solve real-world measurement problems that require unit conversions
determine two similar right triangles when a scale factor is given
make predictions from data tables and graphs to solve problems
plot data on a histogram
calculate the mean and median of a set of data</td>
</tr>
<tr>
<td>AND with Moderate task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
<td>AND with High task complexity –
Common problems presented in mathematical context using various mathematical terms and symbols</td>
</tr>
<tr>
<td>identify the linear representation of a provided real-world situation
use an equation or a linear graphical representation to solve a word problem</td>
<td>identify the linear representation of a provided real-world situation
use an equation or a linear graphical representation to solve a word problem
identify a histogram that represents a provided data set</td>
</tr>
</table>

*Levels 2, 3, and 4 include demonstration of skills described in previous performance levels.

Grade 5 Science Performance Level Descriptors

(for American Samoa, Arizona, BIE, CNMI, Guam, Maine, USVI, and Vermont)

Policy				
Standards	**Level 1 (Beginning – in need of additional support)** *Students at Level 1 are beginning to access the science content and can be expected to need additional support to demonstrate knowledge and skills of the K–12 science framework Extended Performance Expectations.* *Students attempt to perform basic science tasks but will require additional support in order to demonstrate knowledge and skills of the K–12 science framework Extended Performance Expectations by using disciplinary core ideas, practices, and/or crosscutting concepts to address more basic and concrete science phenomena and problems in Level 1.*	**Level 2 (Approaching Expectations)** *Students at Level 2 can be expected to demonstrate developing knowledge and skills of the K–12 science framework Extended Performance Expectations.* *Students can be expected to demonstrate developing knowledge and skills in some disciplinary core ideas together with some aspects of the practices and crosscutting concepts from the Extended Performance Expectations to address primarily basic and concrete science phenomena and problems in Level 2.* *At Level 2, students are expected to have the knowledge and skills of Level 1 and may be able to demonstrate some of the knowledge and skills described in Level 3.*	**Level 3 (Meeting Expectations)** *Students at Level 3 can be expected to demonstrate knowledge and skills of the K–12 science framework Extended Performance Expectations.* *Students can be expected to demonstrate knowledge and skills in the majority of disciplinary core ideas, practices, and crosscutting concepts from the K–12 science framework Extended Performance Expectations to address moderately complex science phenomena and problems, some concrete and some abstract in Level 3.* *At Level 3, students are expected to have the knowledge and skills of Level 2 and may be able to demonstrate some of the knowledge and skills described in Level 4.*	**Level 4 (Exceeding Expectations)** *Students at Level 4 can be expected to demonstrate understanding and skills of the K–12 science framework Extended Performance Expectations.* *Students can be expected to demonstrate understanding and skills in the disciplinary core ideas, practices, and crosscutting concepts from the K–12 science framework Extended Performance Expectations in more sophisticated ways than students in Level 3 to address science phenomena and problems that are complex, more abstract, and/or multi-factorial. Students are expected to describe, explain, and/or respond to phenomena and problems using reasonably complex evidence, analysis, and inference in Level 4.* *At Level 4, students are expected to have the knowledge and skills described in Level 3.*
Range				
PS-1 Matter and Its Interactions • 5-PS1-2 **SEP** • Using Mathematics and Computational Thinking **CCC** • Scale, Proportion, and Quantity	Attempt to identify the appropriate tools or units of measurement (for weight, time, temperature, or volume) for a scientific task.	Identify the appropriate tools or units of measurement (for weight, time, temperature, or volume) for a scientific task.	Compare the weight of matter before and after heating, cooling, or mixing by using data.	Show that the weight of matter does not change when substances are heated, cooled, or mixed by measuring, graphing, or using mathematical relationships.

SEP—Science and Engineering Practices; CCC—Crosscutting Concepts; ESS—Earth and Space Science; PS—Physical Science; LS—Life Science

Grade 5 Science Performance Level Descriptors
(for American Samoa, Arizona, BIE, CNMI, Guam, Maine, USVI, and Vermont)

Range				
PS-2 Motion and Stability: Forces and Interactions • 3-PS2-2 • 5-PS2-1 **SEP** • Planning and Carrying Out Investigations (3-PS2-2, Supporting 5-PS2-1) • Engaging in Argument from Evidence (5-PS2-1) • Analyzing and Interpreting Data (Supporting 3-PS2-2) • Developing and Using Models (Supporting 5-PS2-1) **CCC** • Patterns (3-PS2-2, Supporting 5-PS2-1) • Cause and Effect (5-PS2-1)	Attempt to identify patterns in the motion of an object by using observations or data. Attempt to identify patterns in the motion of falling objects on Earth by using observations.	Identify patterns in the motion of an object by using observations or data. Identify patterns in the motion of falling objects on Earth by using observations.	Predict the future motion of an object by using observations or data. Show the direction objects move when released on Earth (downward toward Earth's center) by identifying or developing a model.	Determine predictable patterns in the motion of an object by describing observations or measurements that can be made in an investigation. Support the claim that Earth's gravity pulls objects downward (toward Earth's center) by describing evidence (observations, data, or a model).
PS-3 Energy • 4-PS3-4 • 5-PS3-1 **SEP** • Constructing Explanations and Designing Solutions (4-PS3-4) • Developing and Using Models (5-PS3-1) **CCC** • Energy and Matter (4-PS3-4, 5-PS3-1) • Patterns (Supporting 5-PS3-1)	Attempt to identify various forms of energy present in a system. Attempt to identify that the Sun is a source of energy for ecosystems.	Identify the various forms of energy involved in energy transfers that occur in an everyday object or device. Identify the Sun as a source of energy for ecosystems by using patterns in food chains or drawings of ecosystems.	Describe the various ways that energy transfer can occur between everyday objects or devices. Describe the direction of energy transfer between two organisms (e.g., plant-animal, animal-animal) or between the Sun and a plant by using a model.	Identify which design or improvement will maximize energy transfer from one form to another by designing or modifying a device. Describe how the energy animals obtain from food comes from the Sun by using a model.

SEP—Science and Engineering Practices; CCC—Crosscutting Concepts; ESS—Earth and Space Science; PS—Physical Science; LS—Life Science

Grade 5 Science Performance Level Descriptors
(for American Samoa, Arizona, BIE, CNMI, Guam, Maine, USVI, and Vermont)

Range				
LS-1 From Molecules to Organisms: Structures and Processes • 4-LS1-1 **SEP** • Engaging in Argument from Evidence • Developing and Using Models (Supporting) • Analyzing and Interpreting Data (Supporting) **CCC** • Systems and System Models • Structure and Function (Supporting)	Attempt to identify the parts of plants or animals that have a specific function by using evidence from data and/or a model.	Identify the parts of plants or animals that have specific functions by using evidence from data and/or a model.	Describe how parts of plants or animals have specific functions that help them survive, grow, or reproduce by using data and/or a model.	Describe evidence to support a claim that parts of plants and/or animals have specific functions that help them survive, grow, or reproduce by using evidence from data and/or a model.
LS-3 Heredity: Inheritance and Variation of Traits • 3-LS3-1 **SEP** • Analyzing and Interpreting Data • Obtaining, Evaluating, and Communicating Information (Supporting) **CCC** • Patterns	Attempt to identify patterns in trait variations between parents and their baby/babies by using data or observations.	Identify patterns in trait variations between parents and their offspring by using data or observations.	Describe patterns in trait variation between groups of organisms (e.g., parents and their offspring, siblings, populations of similar organisms) by using data or observations.	Describe how patterns in trait variation between groups of organisms (e.g., parents and their offspring, siblings, populations of similar organisms) provide evidence of inheritance between parents and their offspring and that there are differences in these traits by analyzing and interpreting data.
LS-4 Biological Evolution: Unity and Diversity • 3-LS4-1 **SEP** • Analyzing and Interpreting Data • Obtaining, Evaluating, and Communicating Information (Supporting) **CCC** • Scale, Proportion, and Quantity	Attempt to recognize that there was life on Earth long ago by using fossils and/or data.	Identify that plants and/or animals lived on Earth long ago by using information about fossils and/or data.	Describe how modern-day plants or animals compare to their ancestors by using observations of fossils and/or data.	Describe the type of environment in which plants and/or animals lived on Earth long ago by using observations of fossils and/or data.

SEP—Science and Engineering Practices; CCC—Crosscutting Concepts; ESS—Earth and Space Science; PS—Physical Science; LS—Life Science

Grade 5 Science Performance Level Descriptors

(for American Samoa, Arizona, BIE, CNMI, Guam, Maine, USVI, and Vermont)

	Range			
ESS-1 Earth's Place in the Universe • 5-ESS1-2 **SEP** • Analyzing and Interpreting Data • Developing and Using Models (Supporting) **CCC** • Patterns • Systems and System Models (Supporting)	Attempt to identify the positions of the Sun, the Moon, and Earth in the solar system by using data or a model.	Identify the positions of the Sun, the Moon, and Earth in the solar system by using data or a model.	Identify patterns concerning the rotation of Earth, Earth's orbit around the Sun, or the Moon's orbit around Earth by analyzing data (e.g., length and direction of shadows, day and night, seasonal appearance of stars) or a model.	Predict or infer patterns concerning the rotation of Earth, Earth's orbit around the Sun, or the Moon's orbit around Earth by analyzing data (e.g., length and direction of shadows, day and night, seasonal appearance of stars) or a model.
ESS-2 Earth's Systems • 3-ESS2-1 • 5-ESS2-1 **SEP** • Analyzing and Interpreting Data (3-ESS2-1) • Planning and Carrying Out Investigations (Supporting 3-ESS2-1) • Developing and Using Models (5-ESS2-1) **CCC** • Patterns (3-ESS2-1) • Systems and Systems Models (5-ESS2-1)	Attempt to describe weather conditions by using observations of weather data. Attempt to identify parts of an Earth system (e.g., geosphere, hydrosphere, atmosphere, biosphere) by using data or a model.	Describe weather conditions by using observations of weather data. Identify parts of an Earth system (e.g., geosphere, hydrosphere, atmosphere, biosphere) by using data or a model.	Describe patterns of weather conditions for a particular season by analyzing weather data. Describe the interaction between two Earth systems (e.g., geosphere, hydrosphere, atmosphere, biosphere) by using a model.	Predict weather conditions for a particular season by analyzing patterns in weather data. Represent the interaction between two Earth systems (e.g., geosphere, hydrosphere, atmosphere, biosphere) by developing a model.
ESS-3 Earth and Human Activity • 5-ESS3-1 **SEP** • Obtaining, Evaluating, and Communicating Information **CCC** • Cause and Effect (Supporting) • Systems and System Models	Attempt to identify a natural or human impact on the environment by using data.	Identify a natural or human impact on the environment by using data.	Describe an effect (positive or negative) of human activities on the environment by using data.	Describe how humans are using science to protect Earth's resources and/or the environment by using data.

SEP—Science and Engineering Practices; CCC—Crosscutting Concepts; ESS—Earth and Space Science; PS—Physical Science; LS—Life Science

Grade 8 Science Performance Level Descriptors

(for American Samoa, Arizona, BIE, CNMI, Guam, Maine, USVI, and Vermont)

Policy				
Standards	**Level 1 (Beginning – in need of additional support)** *Students at Level 1 are beginning to access the science content and can be expected to need additional support to demonstrate knowledge and skills of the K–12 science framework Extended Performance Expectations.* *Students attempt to perform basic tasks but will require additional support in order to demonstrate knowledge and skills of the K–12 science framework Extended Performance Expectations by using disciplinary core ideas, practices, and/or crosscutting concepts to address more basic and concrete science phenomena and problems in Level 1.*	**Level 2 (Approaching Expectations)** *Students at Level 2 can be expected to demonstrate developing knowledge and skills of the K–12 science framework Extended Performance Expectations.* *Students can be expected to demonstrate developing knowledge and skills in some disciplinary core ideas together with some aspects of the practices and crosscutting concepts from the K–12 science framework Extended Performance Expectations to address primarily basic and concrete science phenomena and problems at Level 2.* *At Level 2, students are expected to have the knowledge and skills of Level 1 and may be able to demonstrate some of the knowledge and skills described in Level 3.*	**Level 3 (Meeting Expectations)** *Students at Level 3 can be expected to demonstrate knowledge and skills of the K–12 science framework Extended Performance Expectations.* *Students can be expected to demonstrate knowledge and skills in the majority of disciplinary core ideas, practices, and crosscutting concepts from the K–12 science framework Extended Performance Expectations to address moderately complex science phenomena and problems, some concrete and some abstract at Level 3.* *At Level 3, students are expected to have the knowledge and skills of Level 2 and may be able to demonstrate some of the knowledge and skills described in Level 4.*	**Level 4 (Exceeding Expectations)** *Students at Level 4 can be expected to demonstrate understanding and skills of the K–12 science framework Extended Performance Expectations.* *Students can be expected to demonstrate understanding and skills in the disciplinary core ideas, practices, and crosscutting concepts from the K–12 science framework Extended Performance Expectations in more sophisticated ways than students in Level 3 to address science phenomena and problems that are complex, more abstract, and/or multi-factorial. Students are expected to describe, explain, and/or respond to phenomena and problems using reasonably complex evidence, analysis, and inference at Level 4.* *At Level 4, students are expected to have the knowledge and skills described in Level 3.*
Range				
PS-1 Matter and Its Interactions • MS-PS1-2 **SEP** • Analyzing and Interpreting Data • Planning and Carrying Out Investigations (Supporting) **CCC** • Patterns • Scale, Proportion, and Quantity (Supporting)	Attempt to identify properties of a substance by using data or observations.	Identify properties of a substance by using data or observations.	Determine the identities of substances by using data or observations on the properties of substances.	Determine whether a chemical reaction occurred by using data or observations on the properties of substances before and after an interaction.

SEP—Science and Engineering Practices; CCC—Crosscutting Concepts; ESS—Earth and Space Science; PS—Physical Science; LS—Life Science

Grade 8 Science Performance Level Descriptors

(for American Samoa, Arizona, BIE, CNMI, Guam, Maine, USVI, and Vermont)

Range				
PS-2 Motion and Stability: Forces and Interactions • MS-PS2-2 **SEP** • Planning and Carrying Out Investigations • Analyzing and Interpreting Data (Supporting) **CCC** • Stability and Change • Cause and Effect (Supporting)	Attempt to identify the effects of pushes and pulls on objects by using data from an investigation.	Identify the effects of pushes and pulls on objects by using data from an investigation.	Identify the change in an object's motion when the mass of the object or the force on the object is changed by using data from an investigation.	Describe how the mass of an object or the force on an object will change the motion of the object by using data from an investigation.
PS-3 Energy • MS-PS3-5 **SEP** • Engaging in Argument from Evidence • Asking Questions and Defining Problems (Supporting) • Analyzing and Interpreting Data (Supporting) **CCC** • Energy and Matter	Attempt to determine whether energy is being transferred in a system by asking questions or by using data.	Determine whether energy is being transferred in a system by asking questions or by using data.	Identify the forms of energy that increase or decrease when the kinetic energy of an object changes by using data as evidence.	Make or support a claim that a transfer of energy occurs when the kinetic energy of an object changes by using data as evidence.
PS-4 Waves and Their Applications in Technologies for Information Transfer • MS-PS4-2 **SEP** • Developing and Using Models • Planning and Carrying Out Investigations (Supporting) **CCC** • Structure and Function	Attempt to identify whether a wave is being reflected, absorbed, or transmitted through a material by using data or a model.	Identify whether a wave is being reflected, absorbed, or transmitted through a material by using data or a model.	Describe the path of a wave that is reflected, absorbed, or transmitted through different materials by using a model.	Represent what happens to waves when they are reflected, absorbed, or transmitted through different materials by developing a model.

SEP—Science and Engineering Practices; CCC—Crosscutting Concepts; ESS—Earth and Space Science; PS—Physical Science; LS—Life Science

Grade 8 Science Performance Level Descriptors
(for American Samoa, Arizona, BIE, CNMI, Guam, Maine, USVI, and Vermont)

Range				
LS-1 From Molecules to Organisms: Structures and Processes • MS-LS1-3 • MS-LS1-5 **SEP** • Engaging in Argument from Evidence (MS-LS1-3) • Developing and Using Models (Supporting MS-LS1-3) • Obtaining, Evaluating, and Communicating Information (Supporting MS-LS1-3) • Constructing Explanations and Designing Solutions (MS-LS1-5) • Analyzing and Interpreting Data (Supporting MS-LS1-5) • Asking Questions and Defining Problems (Supporting MS-LS1-5) **CCC** • Systems and System Models (MS-LS1-3) • Cause and Effect (MS-LS1-5)	Attempt to identify structures that are part of human body systems and those that are not by using charts, diagrams, or graphic organizers. Attempt to identify factors that could be affecting the growth of an organism by asking questions.	Identify structures that are part of human body systems and those that are not by using charts, diagrams, or graphic organizers. Identify factors that could be affecting the growth of an organism by asking questions.	Identify those parts that belong to a particular body system and the organization of those parts by using a model. Determine whether a particular factor is affecting the growth of organisms by analyzing data.	Make a claim about two body systems (e.g., circulatory, respiratory, muscular, digestive, nervous, excretory) working together to carry out various functions by using evidence. Explain how the growth of organisms is influenced by various environmental and/or genetic factors by using data.
LS-2 Ecosystems: Interactions, Energy, and Dynamics • MS-LS2-1 • MS-LS2-3 **SEP** • Analyzing and Interpreting Data (MS-LS2-1) • Developing and Using Models (MS-LS2-3) **CCC** • Cause and Effect (MS-LS2-1) • Energy and Matter (MS-LS2-3)	Attempt to identify resources (e.g., food, water, nutrients, space) that are necessary for the growth or survival of organisms or populations of organisms by using data. Attempt to identify the role of organisms (e.g., producer, consumer, decomposer) or nonliving things (e.g., the Sun, water, minerals, air) in cycling energy or matter in an ecosystem by using a model.	Identify resources (e.g., food, water, nutrients, space) that are necessary for the growth or survival of organisms or populations of organisms by using data. Identify the role of organisms (e.g., producer, consumer, decomposer) or nonliving things (e.g., the Sun, water, minerals, air) in cycling energy or matter in an ecosystem by using a model.	Describe the effects of resource availability on organisms and/or populations of organisms by using data or observations. Identify how energy is transferred or that matter is cycled from one specific part of an ecosystem to another specific part by using a model.	Identify evidence of a cause-effect relationship between resource availability and growth of organisms and/or populations of organisms by analyzing data. Describe how energy is transferred or how matter is cycled among living and nonliving parts of ecosystems by developing a model.

SEP—Science and Engineering Practices; CCC—Crosscutting Concepts; ESS—Earth and Space Science; PS—Physical Science; LS—Life Science

Grade 8 Science Performance Level Descriptors

(for American Samoa, Arizona, BIE, CNMI, Guam, Maine, USVI, and Vermont)

Range				
ESS-1 Earth's Place in the Universe • MS-ESS1-1 **SEP** • Developing and Using Models **CCC** • Patterns • Systems and System Models (Supporting)	Attempt to show the positions of Earth (with its tilt), the Sun, and the Moon as Earth orbits the Sun and the Moon orbits Earth in the solar system by identifying a model.	Show the positions of Earth (with its tilt), the Sun, and the Moon as Earth orbits the Sun and the Moon orbits Earth in the solar system by identifying a model.	Describe or compare the positions of the Sun, the Moon, and Earth or the amount or path of light in the cyclic patterns of seasons, lunar phases, or eclipses by using a model.	Compare or show patterns in seasons, lunar phases, or eclipses by using or developing a model of the Earth-Sun-Moon system.
ESS-2 Earth's Systems • MS-ESS2-2 • MS-ESS2-4 **SEP** • Constructing Explanations (MS-ESS2-2) • Obtaining, Evaluating, and Communicating Information (Supporting MS-ESS2-2) • Developing and Using Models (MS-ESS2-4) **CCC** • Scale, Proportion, and Quantity (MS-ESS2-2) • Cause and Effect (Supporting MS-ESS2-2) • Energy and Matter (MS-ESS2-4)	Attempt to identify the process or agent that causes a particular change to Earth's surface by using observations as evidence. Attempt to trace the path of water through Earth's systems by using a model.	Identify the process or agent that causes a particular change to Earth's surface by using observations as evidence. Trace the path of water through Earth's systems by using a model.	Identify whether a geological process or event on Earth was small/large scale and/or whether a process or event happened gradually/rapidly by using information in charts, diagrams, or graphic organizers. Describe the state of water or how water changes state in various parts of the water cycle by using a model.	Explain how geological processes on Earth have caused changes to Earth's surface at various times or spatial scales by using evidence to support an explanation. Describe how the Sun's energy or the force of gravity move water through the water cycle by developing a model.
ESS-3 Earth and Human Activity • MS-ESS3-3 **SEP** • Constructing Explanations and Designing Solutions • Engaging in Argument from Evidence (Supporting MS-ESS3-3) • Asking Questions and Defining Problems (Supporting MS-ESS3-3) **CCC** • Cause and Effect	Attempt to identify an environmental problem caused by human activities/impact by using data.	Identify an environmental problem caused by human activities/impacts by using data.	Make a claim about how a particular method would work to reduce a human impact on the environment by using data.	Select or evaluate a design for a method that is intended to minimize a human impact on the environment by using data.

SEP—Science and Engineering Practices; CCC—Crosscutting Concepts; ESS—Earth and Space Science; PS—Physical Science; LS—Life Science

High School Science Performance Level Descriptors

(for American Samoa, Arizona, BIE, CNMI, Guam, Maine, USVI, and Vermont)

Policy				
Standards	**Level 1 (Beginning – in need of additional support)** *Students at Level 1 are beginning to access the science content and can be expected to need additional support to demonstrate knowledge and skills of the K–12 science framework Extended Performance Expectations.* *Students attempt to perform basic tasks but will require additional support in order to demonstrate knowledge and skills of the K–12 science framework Extended Performance Expectations by using disciplinary core ideas, practices, and/or crosscutting concepts to address more basic and concrete science phenomena and problems in Level 1.*	**Level 2 (Approaching Expectations)** *Students at Level 2 can be expected to demonstrate developing knowledge and skills of the K–12 science framework Extended Performance Expectations.* *Students can be expected to demonstrate developing knowledge and skills in some disciplinary core ideas together with some aspects of the practices and crosscutting concepts from the K–12 science framework Extended Performance Expectations to address primarily basic and concrete science phenomena and problems at Level 2.* *At Level 2, students are expected to have the knowledge and skills of Level 1 and may be able to demonstrate some of the knowledge and skills described in Level 3.*	**Level 3 (Meeting Expectations)** *Students at Level 3 can be expected to demonstrate knowledge and skills of the K–12 science framework Extended Performance Expectations.* *Students can be expected to demonstrate knowledge and skills in the majority of disciplinary core ideas, practices, and crosscutting concepts from the K–12 science framework Extended Performance Expectations to address moderately complex science phenomena and problems, some concrete and some abstract at Level 3.* *At Level 3, students are expected to have the knowledge and skills of Level 2 and may be able to demonstrate some of the knowledge and skills described in Level 4.*	**Level 4 (Exceeding Expectations)** *Students at Level 4 can be expected to demonstrate understanding and skills of the K–12 science framework Extended Performance Expectations.* *Students can be expected to demonstrate understanding and skills in the disciplinary core ideas, practices, and crosscutting concepts from the K–12 science framework Extended Performance Expectations in more sophisticated ways than students in Level 3 to address science phenomena and problems that are complex, more abstract, and/or multi-factorial. Students are expected to describe, explain, and/or respond to phenomena and problems using reasonably complex evidence, analysis, and inference at Level 4.* *At Level 4, students are expected to have the knowledge and skills described in Level 3.*
Range				
PS-1 Matter and Its Interactions • HS-PS1-2 **SEP** • Constructing Explanations and Designing Solutions • Obtaining, Evaluating, and Communicating Information (Supporting) • Developing and Using Models (Supporting) **CCC** • Patterns • Energy and Matter (Supporting)	Attempt to show how substances react in a chemical reaction by using provided information to complete an incomplete chemical reaction model.	Show how substances react by using provided information to complete an incomplete chemical reaction model.	Identify or classify elements that will react similarly in chemical reactions by using a periodic table model.	Construct an explanation for why specific chemical reactions occur by using a periodic table.

SEP—Science and Engineering Practices; CCC—Crosscutting Concepts; ESS—Earth and Space Science; PS—Physical Science; LS—Life Science

High School Science Performance Level Descriptors

(for American Samoa, Arizona, BIE, CNMI, Guam, Maine, USVI, and Vermont)

Range				
PS-2 Motion and Stability: Forces and Interactions • HS-PS2-3 • HS-PS2-5 **SEP** • Constructing Explanations and Designing Solutions (HS-PS2-3) • Engaging in Argument from Evidence (Supporting HS-PS2-3) • Developing and Using Models (Supporting HS-PS2-3) • Planning and Carrying Out Investigations (HS-PS2-5) • Analyzing and Interpreting Data (Supporting HS-PS2-5) **CCC** • Cause and Effect (HS-PS2-3, HS-PS2-5) • Systems and System Models (Supporting HS-PS2-3) • Stability and Change (Supporting HS-PS2-5)	Attempt to identify how forces are acting on a macroscopic object during a collision in a model. Attempt to identify examples of electric current producing magnetic fields or magnetic fields producing electric current by using data or observations.	Identify how forces are acting on a macroscopic object during a collision in a model. Identify examples of electric current producing magnetic fields or magnetic fields producing electric current by using data or observations.	Construct a claim for how a familiar device functions to minimize the forces on a macroscopic object during a collision. Predict or draw conclusions about how a change to a system affects how electric current produces magnetic fields or how magnetic fields produce electric current by using data.	Select, evaluate, or revise the design of a familiar device that minimizes the forces on a macroscopic object during a collision. Plan or conduct an investigation to determine cause-and-effect relationships between magnetic fields and electric current.
PS-3 Energy • HS-PS3-2 **SEP** • Developing and Using Models • Asking Questions and Defining Problems (Supporting) **CCC** • Energy and Matter	Attempt to identify questions that would determine whether an object's kinetic or potential energy is changing in a system.	Identify questions that would determine whether an object's kinetic or potential energy is changing in a system.	Show how kinetic and potential energy change in a system when an object's position changes or when the particles making up an object change their motion by using a model.	Develop or use models to describe how energy is conserved at the macroscopic or particle level when kinetic and/or potential energy are transferred or converted from one form to another in a system.

SEP—Science and Engineering Practices; CCC—Crosscutting Concepts; ESS—Earth and Space Science; PS—Physical Science; LS—Life Science

Range				
LS-2 Ecosystems: Interactions, Energy, and Dynamics • HS-LS2-2 **SEP** • Using Mathematics and Computational Thinking • Obtaining, Evaluating, and Communicating Information (Supporting) **CCC** • Scale, Proportion, and Quantity • Cause and Effect (Supporting)	Attempt to identify factors that affect population size or biodiversity by using provided information.	Identify factors that affect population size or biodiversity by using provided information.	Describe how a factor affects population size or biodiversity in an ecosystem by interpreting data.	Explain how a factor affects population size or biodiversity in an ecosystem at different scales (e.g., habitat size compared to population size) by using mathematical representations of data.
LS-3 Heredity: Inheritance and Variation of Traits • HS-LS3-1 **SEP** • Asking Questions and Defining Problems • Developing and Using Models (Supporting) • Obtaining, Evaluating, and Communicating Information (Supporting) **CCC** • Cause and Effect • Structure and Function (Supporting) • Patterns (Supporting)	Attempt to identify the function of DNA or chromosomes by using provided information.	Identify the function of DNA or chromosomes by using provided information.	Describe how genes and traits are inherited from parents to offspring by using a model.	Ask questions that will provide information about the cause-and-effect relationships among DNA/chromosomes and/or traits that are inherited from parents to offspring.

SEP—Science and Engineering Practices; CCC—Crosscutting Concepts; ESS—Earth and Space Science; PS—Physical Science; LS—Life Science

High School Science Performance Level Descriptors

(for American Samoa, Arizona, BIE, CNMI, Guam, Maine, USVI, and Vermont)

Range				
LS-4 Biological Evolution: Unity and Diversity • HS-LS4-1 • HS-LS4-3 **SEP** • Obtaining, Evaluating, and Communicating Information (HS-LS4-1, Supporting HS-LS4-3) • Analyzing and Interpreting Data (HS-LS4-3, Supporting HS-LS4-1) **CCC** • Patterns (HS-LS4-1, HS-LS4-3) • Stability and Change (Supporting HS-LS4-1)	Attempt to identify how organisms have changed over time by using provided information. Attempt to identify physical traits that can vary in an organism by using provided information.	Identify how organisms have changed over time by using provided information. Identify physical traits that can vary in an organism by using provided information.	Draw conclusions about patterns of relatedness among organisms by using data (e.g., DNA sequences, amino acid sequences, structures found in organisms, embryos, fossils). Describe changes in the distribution of physical traits that can vary in a population by using data.	Describe how comparing patterns in data (e.g., DNA sequences, amino acid sequences, structures found in organisms, embryos, fossils) provide evidence for evolution and common ancestry of living things. Demonstrate that organisms with helpful traits increase in proportion to organisms lacking those traits by using data as evidence.
ESS-1 Earth's Place in the Universe • HS-ESS1-6 **SEP** • Constructing Explanations and Designing Solutions • Asking Questions and Defining Problems (Supporting) • Analyzing and Interpreting Data (Supporting) **CCC** • Stability and Change • Patterns (Supporting)	Attempt to identify patterns in data about ancient Earth materials, meteorites, or other planetary surfaces by using data.	Identify patterns in data about ancient Earth materials, meteorites, or other planetary surfaces by using data.	Describe Earth's formation and early history by asking questions about ancient Earth materials, meteorites, and other planetary surfaces.	Explain Earth's formation and early history by using data about ancient Earth materials, meteorites, or other planetary surfaces.

SEP—Science and Engineering Practices; CCC—Crosscutting Concepts; ESS—Earth and Space Science; PS—Physical Science; LS—Life Science

Range				
ESS-2 Earth's Systems • HS-ESS2-4 • HS-ESS2-5 **SEP** • Developing and Using Models (HS-ESS2-4) • Planning and Carrying Out Investigations (HS-ESS2-5) • Analyzing and Interpreting Data (Supporting HS-ESS2-5) • Asking Questions and Defining Problems (Supporting HS-ESS2-5) **CCC** • Cause and Effect (HS-ESS2-4, Supporting HS-ESS2-5) • Energy and Matter (Supporting HS-ESS2-4) • Structure and Function (HS-ESS2-5)	Attempt to identify how energy flows between two Earth systems by using a model. Attempt to identify testable questions about how water affects Earth's materials and surface processes.	Identify how energy flows between two Earth systems by using a model. Identify testable questions about how water affects Earth's materials and surface processes.	Describe how energy from the Sun drives Earth's climate system by using a model. Use data or observations to draw conclusions about how water affects Earth's materials and surface processes.	Predict or draw conclusions about how various factors (e.g., large volcanic eruptions, human activity, solar output, changes to Earth's orbit and axis, changes to atmospheric composition, etc.) cause changes to Earth's climate (measured as changes in surface temperatures, precipitation patterns, glacial ice volumes, sea levels, biosphere distribution) by using models. Plan or conduct an investigation of the properties of water and its effects on Earth materials and surface processes (e.g., stream transportation and deposition using a stream table, frost wedging by the expansion of water as it freezes, or chemical weathering and recrystallization by testing the solubility of different materials).
ESS-3 Earth and Human Activity • HS-ESS3-4 **SEP** • Constructing Explanations and Designing Solutions • Engaging in Argument from Evidence (Supporting) • Analyzing and Interpreting Data (Supporting) **CCC** • Stability and Change • Cause and Effect (Supporting)	Attempt to identify the impact of positive or negative local human activities on natural systems by using data.	Identify the positive or negative impacts of local human activities on natural systems by using data.	Construct a claim about how a local technological solution reduces the negative impact of human activities on natural systems.	Evaluate or refine the design of a local technological solution that reduces the negative impact of human activities on natural systems.

SEP—Science and Engineering Practices; CCC—Crosscutting Concepts; ESS—Earth and Space Science; PS—Physical Science; LS—Life Science

Appendix C: Scale Score Ranges

Table 1. 2023 Performance-Level Scale Score Ranges by Content Area and Grade

Performance Level	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	High School
English Language Arts							
Level 4	1254–1290	1259–1290	1256–1290	1251–1290	1255–1290	1250–1290	1255–1290
Level 3	1240–1253	1240–1258	1240–1255	1237–1250	1240–1254	1238–1249	1240–1254
Level 2	1234–1239	1234–1239	1232–1239	1231–1236	1236–1239	1230–1237	1236–1239
Level 1	1200–1233	1200–1233	1200–1231	1200–1230	1200–1235	1200–1229	1200–1235
Mathematics							
Level 4	1254–1290	1251–1290	1253–1290	1251–1290	1254–1290	1251–1290	1250–1290
Level 3	1242–1253	1239–1250	1240–1252	1239–1250	1240–1253	1240–1250	1240–1249
Level 2	1235–1241	1232–1238	1232–1239	1233–1238	1234–1239	1234–1239	1235–1239
Level 1	1200–1234	1200–1231	1200–1231	1200–1232	1200–1233	1200–1233	1200–1234

American Samoa, Arizona, BIE, CNMI, Guam, Maine, USVI, and Vermont administered Science in grades 5, 8, and high school. Provisional scale score ranges have been established and will be finalized in summer 2023.

CONFIDENTIAL

Multi-State Alternate Assessment

Name: FIRSTNAME LASTNAME
ID: D133
School: Demonstration School
Test Date: Spring 2023
Grade: 5

What Is In This Report?
Page 1: Contains a summary of your child's performance on this year's test.
Page 2: Contains an introductory letter from MSAA and next steps to support your child.

Performance Summary

FIRSTNAME's performance in English Language Arts and Mathematics is described below.

English Language Arts

Performance Level	Score
Level 3	**1240**

1240

Level 1	Level 2	Level 3	Level 4
1200-1231	1232-1239	1240-1255	1256-1290

$\underbrace{\text{Level 3} \quad \text{Level 4}}_{\text{Meets Expectations}}$

A student's test score can vary. If your child were to be tested again, it is likely that they would receive a score between 1237 and 1243.

English Language Arts consists of Reading and Writing. See below for percent of possible points earned in each area.

Reading 77% **Writing 60%**

Mathematics

Performance Level	Score
Level 3	**1243**

1243

Level 1	Level 2	Level 3	Level 4
1200-1231	1232-1239	1240-1252	1253-1290

$\underbrace{\text{Level 3} \quad \text{Level 4}}_{\text{Meets Expectations}}$

A student's test score can vary. If your child were to be tested again, it is likely that they would receive a score between 1239 and 1247.

Performance Level Descriptors

The scale score and performance level for each content area above summarize FIRSTNAME's performance on the English Language Arts (ELA) and Mathematics tests. The performance level descriptors below describe the knowledge and skills that children who perform at this level generally demonstrate.

English Language Arts

- ✓ use literary texts with clear to implied ideas to compare characters, settings, and events, summarize a text, and use details to answer questions about the text
- ✓ use informational texts with clear to implied ideas to identify the main idea and supporting details, use details to support an author's point, and compare and contrast information and events in different texts
- ✓ use context to define multiple meaning words
- ✓ support an explanatory text topic with relevant information
- ✓ write a narrative with partial command of organization, idea development and/or conventions

Mathematics

- ✓ solve problems with whole numbers, fractions or decimals using mathematical language and symbolic representations (e.g., <, >, =)
- ✓ identify place values
- ✓ round decimals
- ✓ identify the effects of multiplication
- ✓ convert standard measurements including minutes and hours
- ✓ locate a given point on a coordinate plane
- ✓ make comparisons between data sets

 Page 1

Dear Parents and Guardians,

This report summarizes your child's performance on the online 2023 Multi-State Alternate Assessment (MSAA). This report shows the scaled score and performance levels in English Language Arts (ELA) and Mathematics. Also shown is the percent of possible points earned in Reading and Writing. The performance level descriptors describe the knowledge and skills that children who perform at this level generally demonstrate.

The MSAA is designed to assess students in grades 3-8 and High School with significant cognitive disabilities and measures academic content that is aligned to and derived from your state's content standards. The test contains many built-in supports that allow students to take the test using materials they are most familiar with and to communicate what they know and can do. These are some of the built-in supports found in the MSAA:

- shortened ELA reading passages
- pictures, charts, tables, and maps to help students understand the reading passages
- models and examples that explain important ideas and concepts
- smaller numbers on the mathematics tests

To support communication independence to the greatest extent possible, the MSAA is designed to work with different communication modes and systems. Please discuss the supports your child used on the MSAA with your child's teacher.

More information and resources for helping your child are available at your state's alternate assessment web page or by talking with your child's teacher. If you require this letter or your child's report in a different format, please contact your state's department of education.

What skills can be worked on next?

English Language Arts

+ Summarize a text
+ Summarize a text and use inferences
+ Use content vocabulary
+ Use transition words in writing

Mathematics

+ Use mathematical terms and symbols (<, >, =)
+ Solve problems related to percent, rates, and ratios
+ Find the area of a parallelogram
+ Identify numbers on a number line
+ Solve word problems
+ Identify mean, median, and mode
+ Solve equations with decimals

What now?

Bring this report to your next conference with FIRSTNAME's teachers.
You can ask FIRSTNAME's teachers:

- What is FIRSTNAME learning in ELA and Mathematics this year?
- How is FIRSTNAME doing?
- How can I use this information to work with FIRSTNAME this year?
- What resources should I use to support FIRSTNAME?

CONFIDENTIAL

Name: FIRSTNAME LASTNAME
ID: D137
School: Demonstration School
Test Date: Spring 2023
Grade: 5

What Is In This Report?
Page 1: Contains a summary of your child's performance on this year's test.
Page 2: Contains an introductory letter from MSAA and next steps to support your child.

Performance Summary

FIRSTNAME's performance in Science is described below.

Science

Performance Level	Score
Level 3	**1245**

1245

Level 1	Level 2	Level 3	Level 4
1200-1237	1238-1242	1243-1255	1256-1290

Meets Expectations

A student's test score can vary. If your child were to be tested again, it is likely that they would receive a score between 1243 and 1247.

Performance Level Descriptors

The scale score and performance level for the content area above summarizes FIRSTNAME's performance on the Science test. The performance level descriptors below describe the knowledge and skills that children who perform at this level generally demonstrate.

Science

- ✓ Compare the weight of matter before and after a change
- ✓ Predict the future motion of an object
- ✓ Show the direction objects move when released on Earth
- ✓ Describe energy transfer between everyday objects or devices
- ✓ Describe the direction of energy transfer between two organisms
- ✓ Describe how parts of plants or animals have specific functions
- ✓ Describe patterns in trait variations between groups of organisms
- ✓ Describe how modern-day plants or animals compare to their ancestors
- ✓ Identify patterns of rotation or Earth's orbit around the Sun or the Moon's orbit around Earth
- ✓ Describe patterns of weather conditions for a particular season
- ✓ Describe the interaction between two Earth systems
- ✓ Describe an effect of human activities on the environment

Dear Parents and Guardians,

This report summarizes your child's performance on the online 2023 Multi-State Alternate Assessment (MSAA). This report shows the scaled score and performance levels in Science. The performance level descriptors describe the knowledge and skills that children who perform at this level generally demonstrate.

The MSAA Science is designed to assess students in grades 5, 8, and High School with significant cognitive disabilities and measures academic content that is aligned to and derived from your state's science content standards. The test contains many built-in supports that allow students to participate using materials they are most familiar with and to communicate what they know and can do. These are some of the built-in supports found in the MSAA Science:

- pictures, charts, tables, and diagrams to help students understand the science concept
- models and examples that explain important ideas and concepts
- use of concrete science terminology and scenarios

To support communication independence to the greatest extent possible, the MSAA is designed to work with different communication modes and systems. Please discuss the supports your child used on the MSAA with your child's teacher.

More information and resources for helping your child are available at your state's alternate assessment web page or by talking with your child's teacher. If you require this letter or your child's report in a different format, or if you have questions about provisional performance levels and scaled scores, please contact your state's department of education.

What skills can be worked on next?

Science

+ Use charts, graphs, and models to answer questions

+ Focus on physical science concepts such as,
  - the identities of substances
  - motion
  - changes in forms of energy
  - paths of waves

+ Focus on life science concepts such as,
  - body systems
  - organism growth
  - effects of resource availability
  - energy transfer in an ecosystem

+ Focus on Earth and space science concepts such as,
  - positions of the Sun, the Moon, and Earth
  - seasons, lunar phases, and eclipses
  - geological processes or events on Earth
  - the water cycle
  - human impacts on the environment

What now?

Bring this report to your next conference with FIRSTNAME's teachers.
You can ask FIRSTNAME's teachers:
- What is FIRSTNAME learning in Science this year?
- How is FIRSTNAME doing?
- How can I use this information to work with FIRSTNAME this year?
- What resources should I use to support FIRSTNAME?

Multi-State Alternate Assessment

APPENDIX—I

DIFFERENTIAL ITEM FUNCTIONING RESULTS

Table I-1. DIF—Dichotomous

Grade	Group		Low DIF		High DIF	
	Reference	Focal	Favoring		Favoring	
			Reference	Focal	Reference	Focal
5	Male	Female	4	4	0	0
	White	Hispanic or Latino	3	0	0	1
8	Male	Female	2	2	0	0
	White	Hispanic or Latino	1	3	0	0
11	Male	Female	2	1	0	0
	White	Hispanic or Latino	3	3	1	0

Appendix—J
Item Response Theory Parameters

Table J-1. IRT Parameters for Science Grade 5

IREF	a	SE (a)	b	SE (b)
555837	0.59	0.06	-0.23	0.08
555885	0.75	0.06	0.04	0.06
555891	0.33	0.00	-0.27	0.00
556202	0.27	0.00	1.28	0.00
555746	1.29	0.00	-0.40	0.00
555792	1.17	0.09	-0.31	0.05
555834	0.75	0.07	-0.29	0.07
555882	0.36	0.00	1.19	0.00
555889	0.40	0.05	0.61	0.12
555737	0.73	0.00	-0.34	0.00
555849	1.04	0.08	-0.25	0.05
555887	0.35	0.05	0.86	0.15
555951	0.79	0.00	0.00	0.00
555978	0.57	0.06	0.02	0.08
555728	0.73	0.00	-0.22	0.00
555799	0.74	0.07	-0.31	0.07
555872	0.43	0.00	0.29	0.00
555874	0.32	0.05	0.31	0.14
555880	0.37	0.00	0.31	0.00
555894	0.42	0.05	0.54	0.11

IREF	a	SE (a)	b	SE (b)
555957	0.34	0.00	0.67	0.00
556998	0.75	0.00	-0.50	0.00
557040	1.49	0.00	-0.61	0.00
557035	0.96	0.00	-0.35	0.00
557037	0.57	0.06	-0.13	0.08
557429	0.57	0.00	0.20	0.00
556204	0.51	0.00	-0.02	0.00
556982	0.68	0.00	-0.04	0.00
556984	0.48	0.05	0.15	0.09
556996	0.61	0.00	0.17	0.00
557000	0.51	0.05	0.16	0.09
557052	0.84	0.00	-0.16	0.00
555974	0.69	0.06	-0.23	0.07
556986	0.26	0.00	1.29	0.00
557043	0.31	0.00	0.78	0.00
557054	0.77	0.00	0.21	0.00
560422	0.42	0.00	0.61	0.00
560418	0.73	0.00	-0.09	0.00
670555	0.35	0.00	1.44	0.00

Table J-2. IRT Parameters for Science Grade 8

IREF	a	SE (a)	b	SE (b)
558375	0.41	0.05	0.84	0.13
558377	0.55	0.00	-0.17	0.00
558453	0.71	0.06	-0.39	0.07
558475	0.47	0.00	0.58	0.00
558480	0.95	0.08	-0.27	0.05
558366	0.92	0.08	-0.85	0.07
558368	0.78	0.00	-0.47	0.00
558373	0.79	0.07	-0.18	0.06
558451	0.54	0.00	0.44	0.00
558370	0.47	0.00	0.52	0.00
558457	0.22	0.00	1.13	0.00
558469	0.42	0.05	0.41	0.11
558514	0.30	0.04	2.13	0.28
558517	1.29	0.00	-0.52	0.00
558482	0.26	0.00	1.07	0.00
558512	0.40	0.05	0.18	0.11
558526	0.30	0.00	0.79	0.00
558538	0.25	0.04	2.13	0.34
558472	0.32	0.00	0.94	0.00

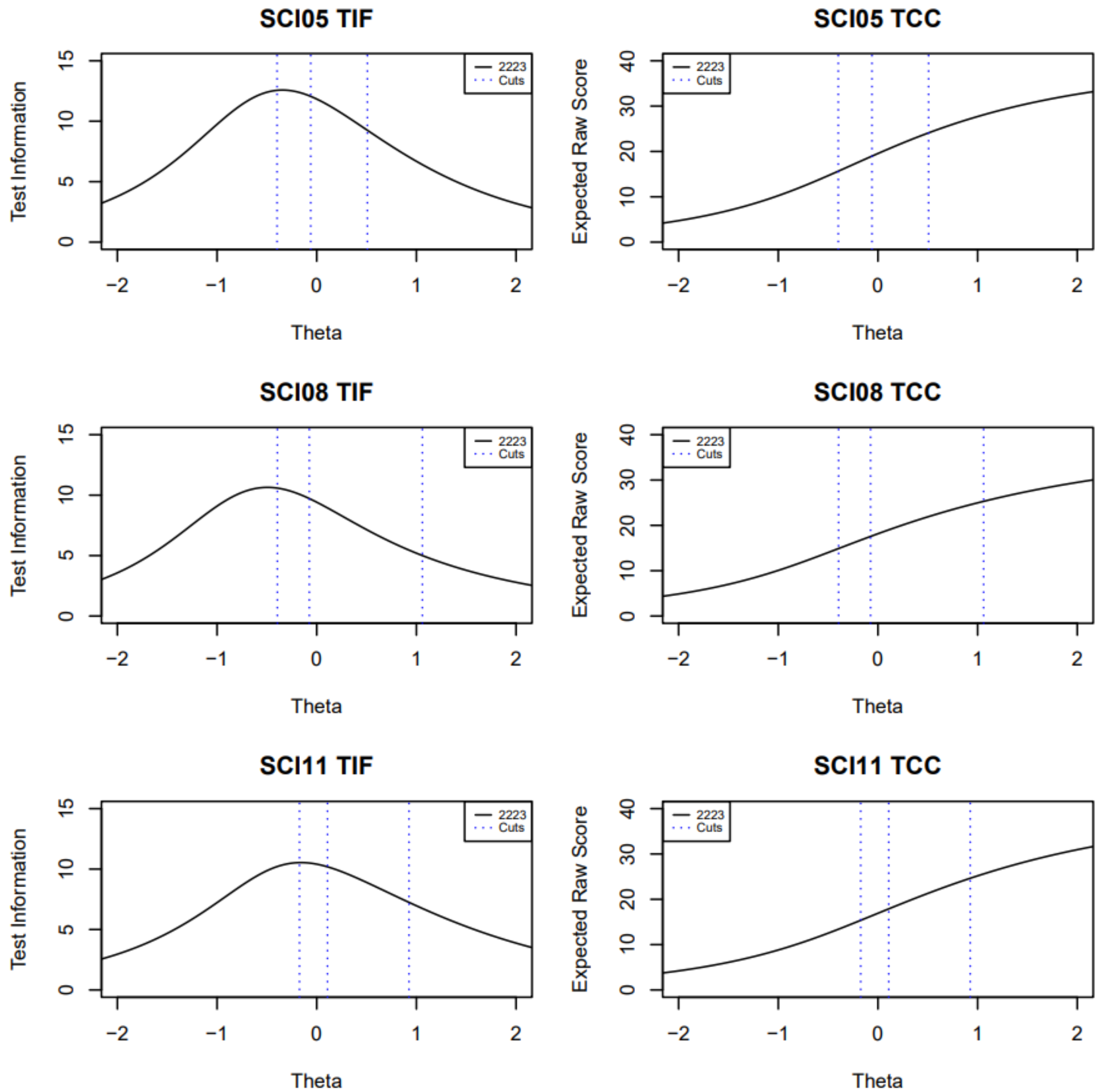
IREF	a	SE (a)	b	SE (b)
558484	0.34	0.00	0.60	0.00
558504	0.55	0.06	-0.30	0.09
572818	1.41	0.00	-0.81	0.00
652251	0.33	0.04	0.08	0.13
558524	0.39	0.00	0.51	0.00
558536	0.70	0.06	-0.04	0.07
558540	0.33	0.00	0.95	0.00
652184	0.85	0.00	-0.47	0.00
572771	0.73	0.00	0.06	0.00
572820	0.34	0.00	0.65	0.00
572822	0.37	0.00	0.33	0.00
652191	0.39	0.00	0.26	0.00
652199	0.54	0.00	0.29	0.00
572767	1.29	0.00	-0.52	0.00
652241	0.33	0.04	1.07	0.17
780917	0.46	0.00	0.16	0.00
780895	1.00	0.08	-0.23	0.05
780906	0.58	0.00	0.34	0.00

Table J-3. IRT Parameters for Science Grade 11

IREF	a	SE (a)	b	SE (b)
555671	1.15	0.10	-0.14	0.05
555701	0.49	0.00	0.62	0.00
555703	0.88	0.08	-0.17	0.06
556348	0.56	0.00	-0.22	0.00
555685	0.84	0.08	-0.06	0.06
555699	0.54	0.00	0.27	0.00
555687	0.48	0.00	0.42	0.00
555689	0.46	0.00	0.44	0.00
555673	0.64	0.07	0.12	0.08
555675	0.38	0.05	0.96	0.17
556352	0.50	0.06	0.38	0.10
556350	0.36	0.00	1.07	0.00
560460	0.98	0.09	-0.46	0.06
560502	0.48	0.00	0.76	0.00
560464	0.32	0.00	0.96	0.00
575331	0.92	0.09	0.02	0.06
781204	0.38	0.00	1.05	0.00
781246	0.68	0.07	0.26	0.08
560506	0.31	0.05	1.08	0.21
560537	0.45	0.06	-0.09	0.11

IREF	a	SE (a)	b	SE (b)
560468	0.58	0.00	0.56	0.00
560496	0.85	0.00	-0.07	0.00
560535	0.81	0.00	-0.08	0.00
575290	0.52	0.00	0.58	0.00
575338	0.56	0.00	0.52	0.00
821123	0.51	0.06	0.07	0.09
821134	0.66	0.07	0.30	0.08
821145	0.27	0.05	1.32	0.27
575296	0.38	0.00	0.62	0.00
560539	0.50	0.00	0.24	0.00
575293	0.44	0.06	1.54	0.21
575335	0.68	0.07	0.18	0.07
781215	0.50	0.06	0.74	0.12
781257	0.34	0.00	0.88	0.00
781270	0.42	0.06	0.73	0.14
781285	0.43	0.00	0.79	0.00
781193	0.64	0.00	0.21	0.00
781298	0.24	0.00	1.59	0.00
781231	1.55	0.13	-0.34	0.04

APPENDIX—K
TEST CHARACTERISTIC CURVES & TEST INFORMATION FUNCTIONS


SCI05 TIF
SCI05 TCC
SCI08 TIF
SCI08 TCC
SCI11 TIF
SCI11 TCC
Test Information
Expected Raw Score
Theta
2223
Cuts
0
5
10
15
0
10
20
30
40
−2
−1
0
1
2

APPENDIX—L
RAW TO SCALED SCORE LOOK-UP TABLES

Table L-1. Raw to Scaled Score Look-up Table—Grade 5 Science

Raw Score	Scaled Score	2021–2022 Standard Error	Performance Level	Scaled Score	2022–2023 Standard Error	Performance Level
0	1200	10.00	1	1200	10.00	1
1	1201	10.00	1	1201	10.00	1
2	1201	10.00	1	1202	10.00	1
3	1208	8.49	1	1209	8.54	1
4	1213	7.02	1	1213	7.06	1
5	1216	6.06	1	1217	6.09	1
6	1219	5.37	1	1220	5.40	1
7	1222	4.87	1	1222	4.90	1
8	1224	4.48	1	1224	4.51	1
9	1226	4.19	1	1226	4.21	1
10	1227	3.96	1	1228	3.98	1
11	1229	3.78	1	1230	3.81	1
12	1230	3.64	1	1231	3.68	1
13	1232	3.54	1	1232	3.58	1
14	1233	3.47	1	1234	3.52	1
15	1234	3.42	1	1235	3.48	1
16	1236	3.40	2	1236	3.46	2
17	1237	3.40	2	1238	3.47	2
18	1238	3.42	2	1239	3.50	2
19	1239	3.46	2	1240	3.54	3
20	1240	3.52	3	1241	3.60	3
21	1242	3.60	3	1243	3.68	3
22	1243	3.70	3	1244	3.77	3
23	1244	3.83	3	1245	3.89	3
24	1246	3.98	3	1246	4.02	3
25	1248	4.15	3	1248	4.18	4
26	1249	4.36	3	1250	4.36	4
27	1251	4.61	3	1252	4.58	4
28	1252	4.89	3	1254	4.83	4
29	1255	5.22	4	1256	5.13	4
30	1257	5.60	4	1258	5.49	4
31	1260	6.05	4	1260	5.92	4
32	1263	6.60	4	1263	6.46	4
33	1266	7.26	4	1267	7.15	4
34	1270	8.09	4	1271	8.05	4
35	1275	9.20	4	1275	9.28	4
36	1281	10.00	4	1282	10.00	4
37	1289	10.00	4	1290	10.00	4
38	1290	10.00	4	1290	10.00	4
39	1290	10.00	4	1290	10.00	4

Table L-2. Raw to Scaled Score Look-up Table—Grade 8 Science

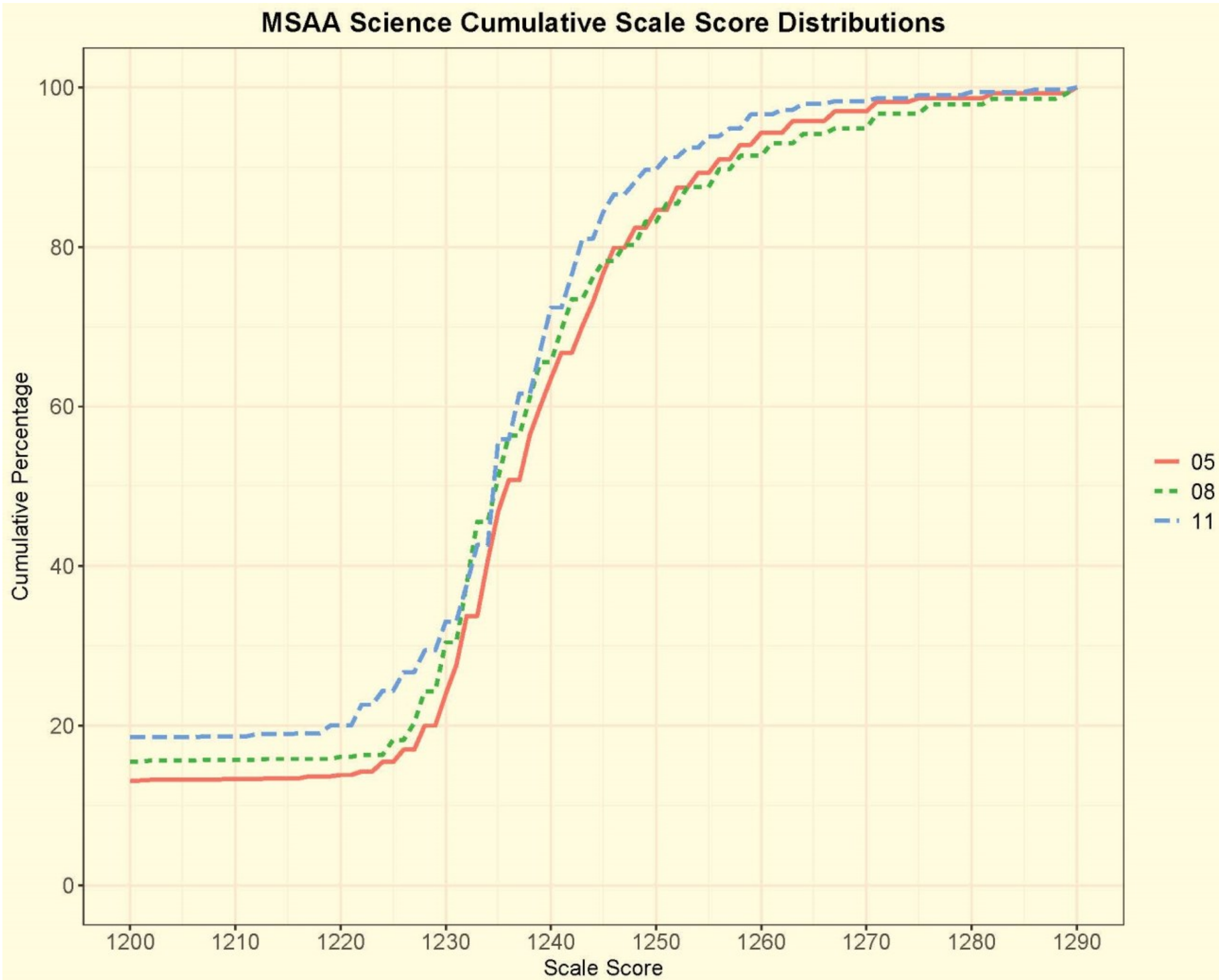
Raw Score	2021–2022 Scaled Score	2021–2022 Standard Error	2021–2022 Performance Level	2022–2023 Scaled Score	2022–2023 Standard Error	2022–2023 Performance Level
0	1200	10.00	1	1200	10.00	1
1	1202	10.00	1	1202	10.00	1
2	1203	10.00	1	1205	10.00	1
3	1205	9.45	1	1207	9.37	1
4	1206	8.89	1	1213	7.58	1
5	1211	7.33	1	1217	6.41	1
6	1215	6.29	1	1220	5.60	1
7	1218	5.56	1	1222	5.03	1
8	1220	5.03	1	1225	4.61	1
9	1222	4.65	1	1227	4.31	1
10	1224	4.38	1	1228	4.09	1
11	1226	4.18	1	1230	3.94	1
12	1228	4.05	1	1232	3.84	1
13	1229	3.96	1	1233	3.78	1
14	1231	3.92	1	1235	3.76	1
15	1232	3.91	1	1236	3.78	2
16	1234	3.94	2	1238	3.82	2
17	1235	4.00	2	1239	3.89	2
18	1237	4.09	2	1241	3.98	3
19	1239	4.21	2	1242	4.10	3
20	1240	4.35	3	1244	4.25	3
21	1242	4.53	3	1245	4.42	3
22	1244	4.74	3	1247	4.62	3
23	1246	4.97	3	1249	4.85	3
24	1247	5.23	3	1251	5.10	3
25	1250	5.53	4	1253	5.40	3
26	1252	5.86	4	1256	5.73	4
27	1255	6.23	4	1258	6.12	4
28	1257	6.64	4	1261	6.56	4
29	1260	7.12	4	1264	7.09	4
30	1263	7.67	4	1267	7.71	4
31	1267	8.31	4	1271	8.48	4
32	1271	9.08	4	1276	9.46	4
33	1276	10.00	4	1282	10.00	4
34	1281	10.00	4	1289	10.00	4
35	1288	10.00	4	1290	10.00	4
36	1290	10.00	4	1290	10.00	4
37	1290	10.00	4	1290	10.00	4

Table L-3. Raw to Scaled Score Look-up Table—Grade 11 Science

Raw Score	2021–2022 Scaled Score	2021–2022 Standard Error	2021–2022 Performance Level	2022–2023 Scaled Score	2022–2023 Standard Error	2022–2023 Performance Level
0	1200	10.00	1	1200	10.00	1
1	1201	10.00	1	1200	10.00	1
2	1203	10.00	1	1200	10.00	1
3	1204	9.66	1	1207	9.11	1
4	1205	9.31	1	1212	7.61	1
5	1210	8.08	1	1216	6.62	1
6	1214	7.19	1	1219	5.91	1
7	1217	6.50	1	1222	5.38	1
8	1220	5.96	1	1224	4.97	1
9	1222	5.52	1	1226	4.65	1
10	1225	5.18	1	1228	4.41	1
11	1227	4.93	1	1230	4.22	1
12	1229	4.75	1	1232	4.08	1
13	1231	4.64	1	1233	3.99	1
14	1232	4.58	1	1235	3.93	1
15	1234	4.57	1	1235	3.90	1
16	1236	4.60	2	1237	3.90	2
17	1238	4.66	2	1239	3.93	2
18	1239	4.73	2	1240	3.97	3
19	1241	4.83	3	1242	4.04	3
20	1243	4.94	3	1243	4.12	3
21	1245	5.06	3	1245	4.22	3
22	1246	5.21	3	1246	4.33	3
23	1249	5.38	4	1248	4.46	3
24	1251	5.58	4	1249	4.61	3
25	1253	5.81	4	1251	4.78	4
26	1255	6.09	4	1253	4.98	4
27	1258	6.42	4	1255	5.19	4
28	1261	6.81	4	1257	5.45	4
29	1264	7.30	4	1259	5.74	4
30	1267	7.90	4	1262	6.08	4
31	1271	8.68	4	1264	6.49	4
32	1276	9.72	4	1267	6.98	4
33	1282	10.00	4	1271	7.61	4
34	1289	10.00	4	1275	8.42	4
35	1290	10.00	4	1280	9.54	4
36	1290	10.00	4	1286	10.00	4
37	1290	10.00	4	1290	10.00	4
38	--	--	--	1290	10.00	4
39	--	--	--	1290	10.00	4

APPENDIX—M
CUMULATIVE SCALE SCORE DISTRIBUTIONS

Figure M-1. Cumulative Scale Score Distributions—Science Grades 5, 8, 11



APPENDIX—N
IRT SUBGROUP RELIABILITY

Note: Values are calculated only for subgroups with 100 or more students.

Table N-1. IRT Subgroup Reliability: Science Grade 5

Description	Number of Students	Scale Score			Standard Deviation	IRT Marginal Reliability	Standard Error
		Minimum	Maximum	Mean			
All	799	1201	1290	1242.93	12.48	0.87	4.29
Female	271	1202	1290	1241.80	11.45	0.86	4.20
Male	525	1201	1290	1243.41	12.91	0.88	4.33
Hispanic or Latino	316	1209	1290	1242.59	11.68	0.86	4.21
American Indian or Alaska Native	41	1224	1263	1241.17	10.09	NA	NA
Asian	24	1224	1290	1236.62	12.96	NA	NA
Black or African American	68	1202	1290	1241.82	13.07	NA	NA
Native Hawaiian or Pacific Islander	15	1231	1263	1241.67	10.18	NA	NA
White (non-Hispanic)	275	1201	1290	1244.15	13.29	0.88	4.44
Two or More Races (non-Hispanic)	42	1224	1290	1243.29	13.57	NA	NA
No Primary race/Ethnicity Undefined	316	1209	1290	1242.59	11.68	0.86	4.21
Currently receiving LEP services	81	1220	1290	1243.99	12.60	NA	NA
Not receiving LEP services	134	1224	1290	1245.26	13.37	0.88	4.47
Economically Disadvantaged Students	86	1224	1290	1246.50	14.12	NA	NA
Non-economically Disadvantaged Students	52	1224	1282	1242.75	11.99	NA	NA
Non-migrant	125	1224	1290	1244.92	13.70	0.88	4.48
Augmentative Communication	175	1209	1275	1237.24	9.73	0.83	3.96
No Augmentative Communication	620	1201	1290	1244.46	12.69	0.87	4.38
Undefined Augmentative Communications	4	1240	1271	1253.50	13.92	NA	NA
Hearing Loss	24	1222	1275	1242.04	12.21	NA	NA
Within Normal Limits	775	1201	1290	1242.95	12.49	0.87	4.29
Visual Impairment	40	1202	1260	1235.42	11.06	NA	NA
Within Normal Limits	757	1201	1290	1243.33	12.44	0.87	4.30
Sensory Stimuli Response	56	1202	1267	1233.70	10.77	NA	NA
Follow Directions	743	1201	1290	1243.62	12.33	0.87	4.30
Special School	59	1217	1271	1236.05	9.05	NA	NA
Regular School Self-contained	432	1201	1290	1241.46	12.11	0.87	4.20
Regular School Primarily Self-contained	154	1213	1290	1243.60	11.86	0.86	4.27
Regular School Resource Room	72	1224	1275	1246.86	11.55	NA	NA
Regular School General Education	82	1226	1290	1250.88	13.69	NA	NA
Communicates Primarily Through Cries	55	1202	1267	1233.24	9.97	NA	NA
Uses Intentional Communication	175	1201	1271	1238.31	10.95	0.85	4.08
Uses Symbolic Language	569	1222	1290	1245.28	12.30	0.86	4.38

Table N-2. IRT Subgroup Reliability: Science Grade 8

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Standard Deviation	IRT Marginal Reliability	Standard Error
All	766	1200	1290	1243.17	14.37	0.88	4.81
Female	295	1200	1290	1242.34	13.29	0.86	4.70
Male	469	1202	1290	1243.65	15.01	0.88	4.88
Hispanic or Latino	298	1202	1290	1243.14	13.57	0.86	4.75
American Indian or Alaska Native	54	1225	1290	1242.13	13.86	NA	NA
Asian	18	1213	1267	1238.83	15.59	NA	NA
Black or African American	58	1225	1289	1243.45	15.88	NA	NA
Native Hawaiian or Pacific Islander	16	1200	1276	1234.31	15.26	NA	NA
White (non-Hispanic)	289	1202	1290	1243.90	14.97	0.88	4.89
Two or More Races (non-Hispanic)	20	1230	1271	1245.20	10.52	NA	NA
No Primary race/Ethnicity Undefined	298	1202	1290	1243.14	13.57	0.86	4.75
Currently receiving LEP services	62	1213	1290	1242.50	15.20	NA	NA
Not receiving LEP services	101	1200	1290	1242.70	15.77	0.89	4.83
Economically Disadvantaged Students	74	1225	1290	1245.28	14.65	NA	NA
Non-economically Disadvantaged Students	33	1222	1290	1243.09	17.00	NA	NA
Non-migrant	89	1222	1290	1244.89	16.25	NA	NA
Augmentative Communication	120	1207	1271	1235.52	9.53	0.79	4.28
No Augmentative Communication	644	1200	1290	1244.55	14.68	0.88	4.91
Undefined Augmentative Communications	26	1225	1276	1241.96	15.12	NA	NA
Hearing Loss	740	1200	1290	1243.21	14.35	0.88	4.81
Within Normal Limits	35	1213	1282	1239.17	15.17	NA	NA
Visual Impairment	729	1200	1290	1243.36	14.31	0.87	4.82
Within Normal Limits	44	1202	1271	1233.00	12.49	NA	NA
Sensory Stimuli Response	722	1200	1290	1243.79	14.25	0.87	4.82
Follow Directions	56	1213	1290	1236.62	12.88	NA	NA
Special School	504	1202	1290	1242.10	13.38	0.87	4.69
Regular School Self-contained	110	1200	1290	1245.66	16.20	0.89	5.13
Regular School Primarily Self-contained	56	1228	1290	1248.46	14.72	NA	NA
Regular School Resource Room	40	1228	1289	1251.50	16.31	NA	NA
Regular School General Education	37	1200	1244	1230.38	9.97	NA	NA
Communicates Primarily Through Cries	149	1207	1289	1239.05	13.88	0.88	4.61
Uses Intentional Communication	580	1220	1290	1245.04	14.11	0.87	4.89
Uses Symbolic Language	766	1200	1290	1243.17	14.37	0.88	4.81

Table N-3. IRT Subgroup Reliability: Science Grade 11

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Standard Deviation	IRT Marginal Reliability	Standard Error
All	672	1200	1290	1240.05	12.14	0.85	4.53
Female	246	1200	1286	1239.95	11.73	0.84	4.51
Male	421	1200	1290	1239.97	12.22	0.85	4.54
Hispanic or Latino	260	1200	1280	1239.15	10.11	0.81	4.37
American Indian or Alaska Native	41	1224	1267	1241.76	10.15	NA	NA
Asian	13	1200	1246	1233.08	11.49	NA	NA
Black or African American	41	1200	1280	1237.39	13.77	NA	NA
Native Hawaiian or Pacific Islander	8	1226	1257	1235.25	9.74	NA	NA
White (non-Hispanic)	264	1200	1290	1241.13	13.30	0.87	4.65
Two or More Races (non-Hispanic)	31	1207	1286	1242.42	14.86	NA	NA
No Primary race/Ethnicity Undefined	260	1200	1280	1239.15	10.11	0.81	4.37
Currently receiving LEP services	29	1224	1262	1237.52	9.51	NA	NA
Not receiving LEP services	108	1200	1290	1241.96	14.21	0.88	4.74
Economically Disadvantaged Students	44	1200	1275	1241.14	12.85	NA	NA
Non-economically Disadvantaged Students	70	1200	1290	1242.27	14.95	NA	NA
Non-migrant	110	1200	1290	1241.81	14.23	0.88	4.74
Augmentative Communication	103	1207	1280	1234.94	10.10	0.80	4.45
No Augmentative Communication	567	1200	1290	1240.97	12.27	0.85	4.55
Undefined Augmentative Communications	18	1200	1246	1232.33	12.25	NA	NA
Hearing Loss	654	1200	1290	1240.26	12.08	0.85	4.53
Within Normal Limits	21	1200	1257	1235.00	12.12	NA	NA
Visual Impairment	648	1200	1290	1240.23	12.14	0.85	4.53
Within Normal Limits	31	1200	1280	1232.97	14.04	NA	NA
Sensory Stimuli Response	641	1200	1290	1240.39	11.95	0.85	4.52
Follow Directions	56	1200	1286	1235.02	14.97	NA	NA
Special School	414	1200	1290	1239.98	11.77	0.85	4.49
Regular School Self-contained	136	1200	1280	1240.32	11.58	0.84	4.52
Regular School Primarily Self-contained	54	1228	1275	1242.56	9.94	NA	NA
Regular School Resource Room	12	1232	1290	1251.50	15.57	NA	NA
Regular School General Education	26	1200	1280	1233.54	13.89	NA	NA
Communicates Primarily Through Cries	137	1200	1290	1235.59	11.57	0.84	4.49
Uses Intentional Communication	509	1200	1290	1241.58	11.81	0.84	4.54
Uses Symbolic Language	672	1200	1290	1240.05	12.14	0.85	4.53

APPENDIX—O
DECISION ACCURACY AND CONSISTENCY RESULTS

Table O-1. Summary of Decision Accuracy (and Consistency) Results by Grade—Overall and Conditional on Performance Level

Grade	DA_Overall	DC_Overall	Kappa	DA_Level1	DA_Level2	DA_Level3	DA_Level4	DC_Level1	DC_Level2	DC_Level3	DC_Level4
5	0.71	0.63	0.50	0.82	0.43	0.57	0.89	0.74	0.32	0.46	0.80
8	0.73	0.66	0.52	0.86	0.37	0.72	0.86	0.77	0.28	0.62	0.78
11	0.71	0.62	0.46	0.80	0.35	0.68	0.84	0.75	0.26	0.56	0.76

Table O-2. Summary of Decision Accuracy (and Consistency) Results by Grade—Conditional on Cutpoint

Grade	DA_Cut1	DC_Cut1	FP_Cut1	FN_Cut1	DA_Cut2	DC_Cut2	FP_Cut2	FN_Cut2	DA_Cut3	DC_Cut3	FP_Cut3	FN_Cut3
5	0.88	0.84	0.05	0.06	0.9	0.85	0.05	0.06	0.91	0.88	0.06	0.03
8	0.88	0.83	0.05	0.07	0.89	0.85	0.05	0.05	0.94	0.92	0.03	0.03
HS	0.85	0.8	0.08	0.06	0.87	0.81	0.06	0.07	0.95	0.92	0.03	0.02

Note: Due to the small sample size, students in Levels 3 and 4 were collapsed for purposes of the decision accuracy and consistency analysis.

APPENDIX—P
LIST OF ACRONYMS

Terms and Acronyms Used in the 2023 MSAA Technical Report	
2PL	two-parameter logistic
AA-AAAS	Alternate Assessment Aligned with Alternate Academic Achievement Standards (current use under ESSA)
AAC	augmentative and alternative communication
AERA	American Educational Research Association
ANOVA	analysis of variance
APA	American Psychological Association
BIE	Bureau of Indian Education
CCC	Crosscutting Concepts
CSEM	conditional standard error of measurement
CTT	Classical Test Theory
DAC	decision accuracy and consistency
DCI	Disciplinary Core Ideas
DETECT	Dimensionality Evaluation to Enumerate Contributing Traits
DIF	differential Item functioning
DIMTEST	computer program used by Cognia
DNU	do not use
DOK	Depth of knowledge
DTA	Directions for Test Administration
ELA	English language arts
EOTS	end-of-test survey
EPE	extended performance expectations
ESEA	Elementary and Secondary Education Act
ESR	early stopping rule
ESSA	Every Student Succeeds Act
FCIP	Fixed Common Item Parameter
HOSS	highest obtainable scale score
HS	High school
ICC	item characteristic curve
ICTC	item category threshold curve
IEP	individualized education program
IIF	Item information function
IRT	Item Response Theory
IT	information technology
KSA	knowledge, skills, and ability
LAL	Links for academic learning
LEP	limited English proficiency
LOSS	lowest obtainable scale score
LI	local independence
LID	local item independence
MSAA	Multi-State Alternate Assessment

continued

Terms and Acronyms Used in the 2023 MSAA Technical Report	
NCME	National Council on Measurement in Education
NCSC	National Center and State Collaborative
NGSS	next generation science standards
PARSCALE	Item response theory (IRT) software program that can perform item analysis and test scoring for dichotomous and polytomous IRT models
PBT	paper-based test
PE	performance expectations
PLAAFP	present level of academic achievement and functional performance?
PLD	performance level descriptor
R9-stringer	student who responds to nine (or more) consecutive multiple-choice items with the exact same option
SCD	Significant cognitive disabilities
SD	standard deviation
SEM	standard error of measurement
SEP	Science and engineering practices
SIU	score interpretations and uses
SQA	Software Quality Assurance
SRC	student response check
TA	test administrators
TC	Test coordinators
TAC	Technical Advisory Committee
TAM	Test Administration Manual
TC	test coordinators
TCC	test characteristic curve
TIF	test information function
UDL	Universal Design for Learning
UWC	use with caution
VI	Volume of Information

APPENDIX Q
ELEMENT LEVEL RATING SCALE

Element Level Rating Scale (from Chapter 11)

11.1 Primary Intended Score Interpretation

The MSAA Science scores provide reliable and valid information about important knowledge and skills in elementary, middle, and high school multidimensional science concepts that students with the most significant cognitive disabilities are attaining.

Assumption 1.1. The content of the test represents the content of the standards (i.e., the Extended Performance Expectations).

The evidence to support this test alignment assumption and its elements was generated in an alignment study that was conducted in 2022. Details regarding the alignment report and evidence of findings is available by contacting the MSAA Science Partners at MSAA@azed.gov.

Element 1.1.1. The Extended Performance Expectations are aligned to the standards (aka, Performance Expectations) from *A Framework for K-12 Science Education* and, as such, aligned to each MSAA Science Partner's academic content standards for each grade level.

Evidence:

The technical manual summarizes evidence from the early development activities that took place. Cognia's development team completed the Performance Expectation prioritization work and then developed the Extended Performance Expectations. Following this, a stakeholder review was conducted that focused on domain coverage, content centrality, performance centrality, and depth of knowledge. Overall, the stakeholders confirmed the domain coverage, content centrality and performance centrality. Feedback provided by stakeholders was incorporated into the Extended Performance Expectations. Study results are listed in Chapter 2.

Relevance: Evidence is highly relevant.

Completeness: Evidence is complete

Overall Support: Existing evidence provides strong support to the element.

Element 1.1.2. The 2023 MSAA Science items are aligned to the Extended Performance Expectations. ***Evidence:*** Item specifications and development and review processes every year are designed and implemented to ensure that items are closely aligned to the Extended Performance Expectations. A new alignment study was conducted in September 2022. Study results are listed in Chapter 3. In summary, the grade level items were well aligned for both content centrality and performance centrality. At least 90% of items were judged as having *some* or *all* of the same performance expectations of the EPEs. Domain concurrence for each grade level was well aligned, at least 90% of the items on the test form align to an EPE defined in the blueprint, no item on the test form reflects expectations not defined in the grade level, and each of the domains in the blueprint is represented by items on the form.

Relevance: Evidence is highly relevant.

Completeness: Evidence is complete

Overall Support: Existing evidence provides strong support to the element.

Element 1.1.3. The MSAA Science Partners have confirmed that the Science Extended Performance Expectations, which are assessed on the MSAA Science, are aligned with each partner's academic content standards for each grade level.

> ***Evidence:*** Membership in MSAA requires each member to adopt the academic content standards that are assessed on the MSAA.

Relevance: Evidence is highly relevant.

Completeness: Evidence is complete.

Overall Support: Existing evidence strongly supports the element.

Possible Challenges: A follow-up study of alignment between the EPEs and each partner's academic content standards may be warranted if the partner's standards are modified.

Element 1.1.4. The 2023 operational MSAA Science items are aligned to the MSAA performance level descriptors.

> ***Evidence:*** The science items are developed following item specifications for each EPE. The item specifications accomplish two purposes: They (1) provide both general and specific guidelines for developing all test items at the grade levels assessed, and (2) describe the test items and provide samples as reference. Sections in the specification documents are dedicated to information about target EPEs, item contexts, variable features, cognitive task levels, use of graphics, item style and format, and general content limits by academic grade-level content target. As such, because items are developed from the item specifications, they are aligned to the performance level descriptors.

Relevance: Evidence is moderately relevant.

Completeness: Evidence is limited.

Overall Support: Existing evidence provides limited support to the element.

Possible Challenges: There are 12 EPEs, with three levels of items, on the blueprint. Each set of PLDs was written specifically for each EPE (12 sets of PLDs in total). With the limitation of the current item bank, MSAA Science does not have items aligned to each EPE and PLD combination (144 combinations). However, this limitation will be alleviated and addressed in future test development cycles.

Assumption 1.2. MSAA Science test items are construct relevant. The elements corresponding to this assumption are concerned with the skills and cognitive processes required to understand and respond to an item in particular and whether they correspond to the skills and processes required in the PLDs.

Element 1.2.1. Items require application of the KSAs of the targeted construct.
Element 1.2.2. Items are accessible for all students.

Element 1.2.3. Appropriate accommodations are provided to meet student needs.
Element 1.2.4. Scaffolding is not a source of construct-irrelevant variance.
Element 1.2.5. Item rendering does not interfere with students' access to test content.
Element 1.2.6. Platform does not interfere with students' ability to interact with test content.
Element 1.2.7. Items are free of bias and sensitivity issues.
The evidence for Elements 1.2.1 through 1.2.7 is interrelated. Some evidence is relevant to a single element. Other evidence is relevant to multiple elements. For that reason, the elements are listed as a group, rather than individually. After listing the evidence for these elements, the relevance of the evidence for each individual element is summarized.

> ***Evidence for 1.2.1:*** The 2023 operational MSAA Science items are aligned to the Extended Performance Expectations, which supports this element. The evidence for Element 1.2.1 is directly linked to Element 1.1.2. As noted above in Element 1.1.2 (Assumption 1.1: The content of the test represents the content of the standards [i.e., the Extended Performance Expectations]), the evidence for 1.1.2 is Complete Evidence. Additionally, in the science standard setting, a process was included in which subject matter experts evaluated the knowledge, skills, and abilities (KSA) demands of the items relative to the KSAs in the PLDs, which provides additional evidence.
>
> ***Evidence for 1.2.1, 1.2.2, 1.2.3, 1.2.4, 1.2.5, 1.2.6, and 1.2.7:*** During the item development process, the items followed a rigorous development cycle, including reviews by MSAA Science Partners and by Item Content and Bias and Sensitivity panelists. See Chapter 3 for a detailed description of the item review process.
>
> ***Evidence for 1.2.7:*** In differential item functioning (DIF) analyses, we examine subgroup differences in performance when sample sizes permit. Actions are taken to ensure that differences in performance are due to construct-relevant, rather than irrelevant, factors. A detailed description of the DIF analysis procedures is given in Chapter 8 along with a summary of the results. Detailed results are presented in Appendix I. Data review also examines the bias and sensitivity of the tested items. Bias and sensitivity checks are also implemented during the item development process.

Summary of evidence for 1.2.1:

Relevance: The evidence is highly relevant.

Completeness: The evidence is complete

Overall Support: The existing evidence strongly supports the claim.

Possible Challenges: A cognitive lab may also help inform the targeted construct. However, since MSAA Science test development uses the experiences from the MSAA ELA and mathematics interaction study, a separate Cognitive Lab is not as critical.

Summary of evidence for 1.2.2:

Relevance: The evidence is highly relevant.

Completeness: The evidence is Moderate to Substantial,

Overall Support: The existing evidence strongly supports the claim.

Possible Challenges: Results of a teacher survey on their experience regarding accessibility during test administration would provide additional evidence.

Summary of evidence for 1.2.3:

Relevance: The evidence is highly relevant.

Completeness: The evidence is Moderate to Substantial.

Overall Support: The existing evidence strongly supports the claim.

Possible Challenges: Results of a teacher survey of their experience regarding accommodations during test administration would provide additional evidence.

Summary of evidence for 1.2.4:

Relevance: The evidence is highly relevant.

Completeness: The evidence is Moderate to Substantial.

Overall Support: The existing evidence strongly supports the claim.

Possible Challenges: A study evaluating whether construct-irrelevant variance is present in items that take new approaches to assessing the EPEs may be warranted.

Summary of evidence for 1.2.5:

Relevance: The evidence is highly relevant.

Completeness: The evidence is Moderate to Substantial.

Overall Support: The existing evidence strongly supports the claim.

Possible Challenges: Results of a teacher survey of their experience regarding any issues having to do with item rendering during test administration would provide additional evidence.

Summary of evidence for 1.2.6:

Relevance: The evidence is highly relevant.

Completeness: The evidence is Moderate to Substantial.

Overall Support: The existing evidence strongly supports the claim.

Possible Challenges: Results of a teacher survey of their experience regarding any issues having to do with the platform during test administration would provide additional evidence.

Summary of evidence for 1.2.7:

Relevance: The evidence is highly relevant.

Completeness: The evidence is Moderate to Substantial.

Overall Support: The existing evidence strongly supports the claim.

Possible Challenges: None

Assumption 1.3. Test administrations in MSAA states in 2023 followed prescribed, standardized procedural requirements.

Element 1.3.1. Test Administrators and School and District Coordinators understood and performed their roles appropriately.

Evidence: Test Administrators participated in mandatory test administration training. Chapter 5, Training and Administration, provides detailed evidence in regard to ensuring the Test Administrators and Test Coordinators properly understood and performed their roles.

Six online training modules address the specific responsibilities of the Test Administrators and provide information from the three documents they were required to use: Test Administrator Manual (TAM), the Directions for Test Administration (DTA), and the MSAA Online Assessment System User Guide for Test Administrators. Additionally, there is a training module specific to science that outlines the background and design of that content area. After completing the training modules, Test Administrators were required to successfully complete a final quiz with a score of 80% or better.

Required training for Test Coordinators. Six online training modules address the responsibilities of the Test Coordinators. Test Coordinators were also provided with the following supporting documents: TAM, DTAs, the MSAA Online Assessment System User Guide for Test Administrators, and the MSAA Online Assessment System User Guide for Test Coordinators. In addition, each Test Administrator:

- received four best practice videos;
- received a technical support chart that provides examples of when and who to contact to obtain answers concerning MSAA assessment or administration; and
- completed a survey. (Results are evidence that address this element.)

All the above evidence is described in detail in Chapter 5.

Observers were sent into the field to observe test administration and complete an observation checklist. The checklists and any accompanying notes provide evidence as to whether the training was effectively followed by the Test Administrators and Test Coordinators.

The summarized results from 2021–22 MSAA Science administration observations are included below:

- Of 10 Arizona observations, 10 administered the MSAA following the instructions in the Directions for Test Administration (DTA).
- All use of DTA was indicated as high fidelity.
- All 10 observations also indicated secure storage of secure test materials.

Relevance: The evidence is highly relevant.

Completeness: The evidence is limited.

Overall Support: The evidence provides limited support for the claim.

Possible Challenges: Additional observations are required for a representative sample.

Element 1.3.2. Test security concerns were limited.

Chapter 5, Section 5.6.11, *Test Security and Test Irregularities*, provides detailed evidence indicating that test security policies and practices resulted in limited test security concerns.

Evidence: Evidence for 1.3.2 includes the following:

- Irregularity reports, which Test Administrators and District Test Coordinators file if disruptions to orderly test administrations occur or if they observe suspicious activity related to test content security or student test data integrity, indicate no significant problems. Specifically, for 2021–22 administration, no instance of irregularity was reported.

Relevance: The evidence is highly relevant

Completeness: The evidence is limited

Overall Support: The evidence provides limited support to the claim

Possible Challenges: None

Assumption 1.4. Test scores on the 2023 MSAA Science provide reliable information about student performance and accurate classifications into provisional performance levels.

Element 1.4.1. MSAA scores and categorizations into provisional performance levels are adequately reliable for their intended purpose.

Evidence: Evidence for 1.4.1 includes the following:

Internal consistency: Chapter 10 provides a description of reliability theory and interpretation, a review of the relevant equations, and a summary of the results. In particular, the reliability estimates can be interpreted as the correlation that would be obtained between scaled scores on two parallel forms. All MSAA tests show adequate reliability IRT marginal reliability value of 0.8 or higher.

Scaled score standard errors: Chapter 9 provides a description of calculation and interpretation of the scaled scores, as well as a description of the calculation of the standard error for a scaled score. The average standard error for a reported scaled score is reported in Chapter 10. 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.

Performance level classification consistency and accuracy estimates: Accuracy is an estimate of the probability that the observed classification is the true classification. Consistency is an estimate of the probability that students would receive the same classification if they tested twice on parallel forms. Chapter 10 describes the theory and equations underlying the estimation of classification accuracy and consistency, while also reporting summary statistics. Detailed results are provided in Appendix O.

Relevance: Evidence is highly relevant.

Completeness: Evidence is complete.

Overall Support: Existing evidence strongly supports the element.

Possible Challenges:

1. Readers could discuss any possible impacts on the reliability evidence that could stem from the local item dependence detected in the dimensionality analysis (see section 8.3). Such a challenge would be balanced against the strong evidence of the unidimensional item response functions fitting the data well, along with the research of Roussos, Li, & Lonczak (2013), who have shown that when total test score is used for ability estimation, there is only negligible under-estimation of the standard errors of the test scores, even when the LID is large. This stems from the fact that when positive LID occurs for some item pairs it is essentially equally balanced by negative LID, the presence of which has been ignored in papers discussing the effects of positive LID (Habing & Roussos, 2003).

Element 1.4.2. Item characteristics support intended interpretations about all students who take the MSAA.

Evidence: The psychometric characteristics most pertinent to evaluating the adequacy of individual items are the estimated item parameters. The item parameter estimates are summarized in tables in Chapter 9. For dichotomously scored items, the item parameters include the discrimination parameter and the difficulty level parameter. All items undergo statistical analyses at the time of field testing, including classical, DIF, and IRT analyses. The results of these analyses are reviewed in Data Review meetings with the MSAA Science Partners. Only those field-tested items that pass statistical and Data Review criteria are passed forward for operational use.

After field-testing and prior to operational administration, items from the previous operational administration will be 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 how the items correlate with each other in terms of the underlying constructs of the test.

Relevance: Evidence is highly relevant.

Completeness: Evidence is complete.

Overall Support: Existing evidence strongly supports the element.

Possible Challenges: Could have discussion of any possible impacts on the ICC and IIF evidence that could stem from the local item dependence detected in the dimensionality analysis. In this regard, such a challenge would be balanced against the strong evidence of the unidimensional item response functions fitting the data well, the paucity of items flagged for DIF, and the research of Roussos, Li, & Lonczak (2013), who have shown that when total test score is used for ability estimation, there is only negligible underestimation of the standard error of the test scores, even when the LID is large. This stems from the fact that when positive LID occurs for some item pairs it is essentially equally balanced by negative LID, the presence of which has been ignored in papers discussing the effects of positive LID (Habing & Roussos, 2003).

Element 1.4.3. Test characteristics support intended interpretations about all students who take the MSAA.

Evidence: Evidence for 1.4.3 includes the following:

- Dimensionality: Dimensionality: Dimensionality analysis was conducted on each grade-level test. Section 8.2 gives a detailed description of the hypothesis testing and effect size estimation methods. Results are summarized in a table accompanied by a description of the results. Grade 8 and high school showed a moderate level of Local Item Dependence (LID). Grade 5 shows a moderate to strong value of LID. The violations seem to be clearly related to examinee stringer behavior and not to any nuisance dimensions that could distort the unidimensional model. As described in the dimensionality section, any effects due to stringers are controlled by carefully limiting the number of items having the same key along any one item set.
- Test Information Functions: Chapter 9 provides a detailed description of the psychometric model that was fitted to the data. In particular, it describes the test information function (TIF), the most pertinent product of the psychometric model, in regard to evaluating the adequacy of the test. Appendix K shows the TIF graphs for all three grades of the MSAA Science. By examining the value of TIF at the performance level cuts (given in the graphs), the psychometric appropriateness of each test can be evaluated. As is evident in the TIFs in all grades, psychometric information is highest around the performance levels

1/2 and 2/3 cut scores. For all science grades, the TIF at these two cuts are either above or approaching the desire benchmark. This means that measurement precision and classification accuracy are maximized in the area of the scale where these cut scores are located and psychometric information about the students whose test performance locates them in that same area also is maximized. This targeting of information around the performance levels 1/2 and 2/3 cut scores is by design.

Relevance: Evidence is highly relevant.

Completeness: Evidence completeness is Moderate to Substantial.

Overall Support: Moderately supports the claim. The TIFs indicate that information functions are maximized at two different locations on the theta scale (i.e., the cut scores).

Possible Challenges: Cut 3 TIF is generally below the benchmark, this will be addressed by increasing the item pool each year.

Element 1.4.4. Scaling of the MSAA supports intended interpretations about all students who take the MSAA.

Evidence: Evidence for 1.4.4 includes the following:

Differential item functioning (DIF) analyses: The scale used for reporting scores is assumed to be measuring only those constructs that are intended to be measured by each test. DIF analyses were conducted to detect items that may be measuring construct-irrelevant variance. Subgroup differences in item-level performance are examined when sample sizes permit. If an item is flagged, appropriate actions are taken to investigate whether the differences in performance are due to construct-irrelevant factors. A detailed description of the DIF analysis procedures is given in Chapter 8, along with a summary of the results. Detailed results are presented in Appendix I.

Dimensionality: The scale used for reporting scores is a unidimensional scale. Dimensionality analysis was conducted for each grade-level test to examine the degree to which unidimensionality is evident. When the null hypothesis of unidimensionality is rejected, the dimensionality analysis quantifies the violation of unidimensionality and attempts to describe what may be causing the violation. Section 8.2 gives a detailed description of the hypothesis testing and effect size estimation methods. Results are summarized in a table accompanied by a description of the results. Mostly small to moderate violations of local independence were noted, and interpretations of these results were presented. The moderate violations and the few strong violations of local independence seem to be clearly related to aberrant student behavior (stringer effects), rather than to any nuisance dimensions.

Calibration: The unidimensional scale used for reporting scores is based on an underlying unidimensional IRT model. The initial form of the IRT model is established by an initial calibration of the item response data. The calibration must be conducted accurately in order for the scaling to be appropriately implemented. Section 9.2 provides evidence that can be used to evaluate the effectiveness of the calibration. The evidence provided for the calibration procedure includes discussion of the removal of stringers and a description of how the convergence of the statistical calibration was evaluated.

Model fit: After the initial calibration converged to a solution, the fit of the model was evaluated. Section 9.2 described how model fit was evaluated and the criteria that were used.

Relevance: The evidence is highly relevant.

Completeness: The evidence is complete.

Overall Support: Moderate to Strong.

Possible Challenges: Moderate to strong violations of local independence are a challenge to the label of strong support, but this challenge is counterbalance by the many analyses indicating goodness of fit of the unidimensional ICC's as well as a lack of any evidence that the strong violations are related to nuisance dimensions. In addition, the use of total score as the basis for scaling has been shown by research (Roussos, Li, & Lonczak; 2013) to result in only negligible underestimation of ability standard errors. Furthermore, an article by Ip (2010) demonstrates the empirical indistinguishability of multidimensional IRT and locally dependent unidimensional IRT models.

Element 1.4.5. Equating of MSAA test forms supports intended interpretations about MSAA students.

Evidence: Regarding the equating of MSAA test forms to maintain consistency and support the intended interpretations of MSAA students, it's important to note that, as this year represents another standard setting cycle for the assessment, there was no requirement for equating work. Consequently, no pertinent evidence is provided for this element in the current context.

However, it's worth highlighting that item calibration for the 2023 assessment employed the same well-established IRT scale that was put in place during the previous year.

Relevance: NA

Completeness: Evidence is moderately complete. NA

Overall Support: NA

Possible Challenges: NA

Assumption 1.5. Item and test scoring in 2023 were implemented accurately.

Element 1.5.1. Machine-scored items were scored accurately.

Evidence: Machine-scorable MSAA items are submitted to a key verification process. As mentioned in Chapter 6, all the operational multiple-choice items are examined prior to score reporting to ensure that the option that was designated as the key was indeed the correct response.

Relevance: The evidence is highly relevant.

Completeness: Complete evidence.

Overall Support: Evidence strongly supports the claim.

Possible Challenges: None.

Assumption 1.6. MSAA Science scores correlate as expected with external indicators of student proficiency (i.e., concurrent evidence).

Element 1.6.1. MSAA Science scores correlate as expected with other measures of student proficiency.

Evidence: Peer reviewers acknowledge the challenge of acquiring additional evidence of student achievement that can be correlated with state alternate assessment scores, which they require for state grade-level assessments. As an alternative, peer reviewers do accept correlations that are internal to an alternate assessment as evidence in support of this assumption. (D. Peasley, personal communication to S. Ferrara, October 21, 2019.) The correlations between 2023 MSAA Science scale scores with the ELA and mathematics scores in grades 5, 8, and HS are listed in the table below.

	Correlations	
Grade	**ELA and Science**	**Math and Science**
5	0.87	0.83
8	0.89	0.81
HS	0.85	0.82

The strong positive values between the MSAA Science and the ELA and mathematics scale scores provide convergent validity evidence in the sense that they suggest that students' general academic and communicative capabilities are reflected strongly in both their MSAA Science, and their MSAA ELA and mathematics performances and scores.

Relevance: The evidence is moderately relevant.

Completeness: Limited Evidence.

Overall Support: Evidence provides limited support of the element.

Possible Challenges: The lack of external assessment to correlate with MSAA Science is a possible challenge.

11.2 Primary Intended Score Uses

11.2.1 Primary Intended Score Use 1

Schools and districts use the MSAA and its results to (a) monitor trends in student performance and (b) design professional development for teachers.

Assumption 2.1. MSAA Science scores enable teachers and school, district, and state leaders to monitor trends in student proficiency.

Element 2.1.1. MSAA Science scale scores for groups of students are adequately reliable and valid to enable school, district, and state leaders to monitor changes in means, standard deviations, and proficiency 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 Assumptions 1.4 and 1.5 and in Chapters 6, Scoring, and 10, Reliability. Specifically:

- Evidence of individual score reliability in Section 10.1, IRT Marginal Reliability, is comparable to industry standards for grade-level educational achievement tests. The reliability of aggregated scores (e.g., means) usually is as high as or higher than individual score reliabilities (Brennan, 1995).
- Evidence presented in Appendix N and discussed in Section 10.2, Subgroup Reliability, indicates that reliability for some subgroups is strong. However, Section 10.2 discusses caveats in interpreting subgroup score reliability with caution because of the potential deleterious effects of small subgroup sizes on estimator standard error as well as possible severe restriction of range.
- No other aggregate score reliabilities (e.g., at the school level) exist.
- MSAA survey will be sent out to MSAA Science stakeholders to gauge their perceptions of using the MSAA results to monitor student performance as well as designing professional development.

Relevance: The evidence is moderately relevant.

Completeness: Limited Evidence.

Overall Support: Evidence moderately supports the claim.

Possible Challenges: Possible additional evidence is discussed under Assumptions 1.4, 1.5, and 1.6 Additional information will be collected a few years into the MSAA Science testing cycle.

Element 2.1.2. MSAA Science 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 proficiency level categorizations is presented above under Assumption 1.4. The most pertinent evidence follows.

Performance level classification consistency and accuracy estimates: Accuracy is an estimate of the probability that the observed classification is the true classification. Consistency is an estimate of the probability that students would receive the same classification if they tested twice on parallel forms. Section 10.4 describes the theory and equations underlying the estimation of classification accuracy and consistency, while also reporting summary statistics. Detailed results are provided in Appendix O.

Relevance: The evidence is highly relevant.

Completeness: Complete Evidence.

Overall Support: The evidence strongly supports the element.

Possible Challenges: Additional survey information will be collected a few years into the MSAA Science testing cycle.

Element 2.1.3. The relationship between MSAA Science scores and external measures of student achievement and growth is as expected, compared to grade-level assessments and other measures of student achievement.

Evidence: The internal correlations between 2023 MSAA Science with ELA and mathematics scores for grades 5, 8, and HS are listed in the correlation table in Element 1.6.1. These correlations indicate a moderate to strong relationship between MSAA Science scores with ELA and mathematics scores, which is reasonably consistent with correlations observed between grade-level state assessments and external measures (e.g., local interim assessments). They suggest that MSAA Science scores enable teachers and school, district, and state leaders to monitor trends in student achievement as when, for example, student achievement in both content areas either progress similarly, or do not progress similarly.

Internal correlations are accepted as evidence for critical element 3.4, specifically for alternate assessments, because of the difficulties in collecting additional, external assessment evidence on students with significant cognitive disabilities (D. Peasley, personal communication to S. Ferrara, October 17, 2019).

Relevance: The evidence is highly relevant.

Completeness: Moderate to Substantial Evidence.

Overall Support: Evidence moderately supports the element.

Possible Challenges: Collecting additional evidence to correlate MSAA Science with external assessment.

Assumption 2.2. MSAA Science results are used to design professional development for teachers.

States offer guidance to local districts for developing teacher professional development, as exemplified by the NCSC's document titled "How to Teach the State Standards to Students Who Take Alternate Assessments" (accessible at https://cms.azed.gov/home/GetDocumentFile?id=5866dbe1aadebe085c4de5b4).

Furthermore, in the survey referenced in Assumption 2.1, there is additional evidence regarding the utilization of MSAA results in shaping professional development for educators from the pilot validity survey. It was found that only 40% of pilot survey participants reported offering professional development (PD) opportunities to teachers specifically focused on interpreting and applying MSAA scores. These PD sessions primarily served the purposes of aiding in the identification of Individualized Education Programs (IEPs) and the establishment of performance benchmarks.

Additionally, one-third of the respondents indicated that they conducted MSAA-related presentations, typically on an annual basis. These presentations were primarily targeted at teachers and school/district leaders.

Relevance: Evidence is highly relevant

Completeness: Evidence is moderate to substantial

Overall Support: Existing evidence moderately supports the assumption. It's noteworthy that the demographics of the pilot survey may not entirely reflect the broader MSAA population.

11.2.2 Primary Intended Score Use 2

The MSAA and its results are used to help teachers integrate MSAA scores and other information into their instructional planning.

Assumption 3.1. Teachers use the MSAA Science and its results to better integrate assessment with their instructional planning.

Element 3.1.1. Teachers find the performance level descriptors and their students' performance levels useful for planning instruction, especially students in performance levels 1 and 2.

> ***Evidence:*** Annual compliance monitoring of IEPs in all states indicates that special education teachers refer to PLDs to establish present levels of performance and to inform goals. For example, the Arizona Department of Education guidance on IEP-required components requires that "The IEP includes measurable annual goals, including academic and functional goals that reflect the needs identified in the PLAAFP and current assessment data" (p. D40; see https://cms.azed.gov/home/GetDocumentFile?id=5b2a897d1dcb250f1c55e5b3).

Relevance: The evidence is moderately relevant.

Completeness: Limited Evidence.

Overall Support: The evidence moderately supports the element.

Possible Challenges: The current evidence only support the year 1 use of the assessment. Additional Evidence needed; an example of additional evidence would be a survey of teachers to begin to understand the degree to which teachers find MSAA Science scores and other MSAA-based information useful for planning instruction.

Element 3.1.2. Teachers find their students' scale score information useful for planning instruction, especially students in Levels 1 and 2.

> ***Evidence:*** Annual compliance monitoring of IEPs in all states suggests that special education teachers refer to PLDs to establish present levels of performance and to inform goals. For example, the Arizona Department of Education guidance on IEP required components requires that "The IEP includes measurable annual goals, including academic and functional goals that reflect the needs identified in the PLAAFP and current assessment data" (p. D40; see https://cms.azed.gov/home/GetDocumentFile?id=5b2a897d1dcb250f1c55e5b3).

Relevance: The evidence is moderately relevant.

Completeness: Limited Evidence.

Overall Support: The evidence moderately supports the element.

Possible Challenges: Additional Evidence needed; an example of additional evidence could be a survey of teachers to begin to understand the degree to which teachers find MSAA Science scores useful for planning instruction.

Assumption 3.2. Teachers use MSAA Science scores and other information for instructional planning.

Element 3.2.1. Teachers use MSAA Science scores and other information for planning instruction.

> ***Evidence:*** Annual compliance monitoring of IEPs in all states indicates that special education teachers refer to PLDs to establish present levels of performance and to inform goals. For example, the Arizona Department of Education guidance on IEP required components requires that "The IEP includes measurable annual goals, including academic and functional goals that reflect the needs identified in the PLAAFP and current assessment data" (p. D40; see https://cms.azed.gov/home/GetDocumentFile?id=5b2a897d1dcb250f1c55e5b3).

Relevance: The evidence is moderately relevant.

Completeness: Limited Evidence.

Overall Support: The evidence moderately supports the element.

Possible Challenges: Additional Evidence needed; an example of additional evidence could be a survey of teachers to begin to understand the degree to which teachers use MSAA Science scores and other MSAA-based information for planning instruction.

11.2.3 Primary Intended Score Use 3

Parents use the MSAA and its results to get information about (a) what their child knows and can do, and (b) their child's progress from year to year.

Assumption 4.1. Parents find MSAA Science scores and other information useful for understanding what their child knows and can do.

Element 4.1.1. Parents understand and correctly interpret MSAA Science scores and other information to understand what their child knows and can do.

> ***Evidence:*** For this first year of the assessment, MSAA provides information to guide parents in interpreting and using MSAA scores and other information about their child's achievement and learning needs. For example, the Arizona Department of Education sends to districts a Parent Overview to accompany each child's Individual Score Report. The overviews are available online in both English and Spanish (see https://www.azed.gov/assessment/parents). Similarly, the Maine Department of Education provides the Parent Overview of the MSAA Assessment System (see https://www.maine.gov/doe/sites/maine.gov.doe/files/inline-files/2016ParentOverview-allgradescombined.pdf).

Relevance: The evidence is moderately relevant.

Completeness: Limited Evidence.

Overall Support: The evidence moderately supports the element,

Possible Challenges: Currently, the feedback from parents regarding the score interpretation and clarity of the score reports are lacking. an example of additional evidence could be a survey of parents to begin to understand the degree to which parents correctly understand and interpret MSAA Science scores and other MSAA-based information to understand what their child knows and can do.

Element 4.1.2. Parents use MSAA Science scores and other information appropriately to understand what their child knows and can do and make decisions about their child's education and learning needs.

> ***Evidence:*** MSAA provides information to guide parents in interpreting and using MSAA Science scores and other information about their child's achievement and learning needs. For example, the Arizona Department of Education sends to districts a Parent Overview to accompany each child's Individual Score Report. The overviews are available online in both English and Spanish (see https://www.azed.gov/assessment/parents). Similarly, the Maine Department of Education provides the Parent Overview of the MSAA Assessment System (see https://www.maine.gov/doe/sites/maine.gov.doe/files/inline-files/2016ParentOverview-allgradescombined.pdf).

Relevance: The evidence is moderately relevant.

Completeness: Limited Evidence.

Overall Support: The evidence moderately supports the element

Possible Challenges: Additional evidence needed; An example of additional evidence could be a survey of parents to begin to understand the degree to which parents use MSAA Science scores and other MSAA-based information to understand what their child knows and can do.

Assumption 4.2. Parents find MSAA Science scores and other information useful for understanding their child's progress across grades.

Element 4.2.1. Parents understand and correctly interpret MSAA Science scores and other information to understand their child's progress across grades.

> ***Evidence:*** For the very first year of MSAA Science results, MSAA provides information to guide parents in interpreting and using MSAA Science scores and other information about their child's achievement and learning needs. For example, the Arizona Department of Education sends to districts a Parent Overview to accompany each child's Individual Score Report. The overviews

are available online in both English and Spanish (see https://www.azed.gov/assessment/parents). Similarly, the Maine Department of Education provides the Parent Overview of the MSAA Assessment System (see https://www.maine.gov/doe/sites/maine.gov.doe/files/inline-files/2016ParentOverview-allgradescombined.pdf).

Relevance: The evidence is moderately relevant.

Completeness: Limited Evidence.

Overall Support: The evidence moderately supports the element.

Possible Challenges: This is the first year of the assessment, Additional evidence needed; an example of additional evidence could be a survey of parents to begin to understand the degree to which parents correctly understand and interpret MSAA Science scores and other MSAA-based information to understand their child's progress from year to year**.**

Element 4.2.2. Parents use MSAA Science scores and other information appropriately to understand their child's progress across grades and make decisions about their child's education and learning needs.

Evidence: MSAA provides information to guide parents in interpreting and using MSAA Science scores and other information about their child's achievement and learning needs. For example, the Arizona Department of Education sends to districts a Parent Overview to accompany each child's Individual Score Report. The overviews are available online in both English and Spanish (see https://www.azed.gov/assessment/parents). Similarly, the Maine Department of Education provides the Parent Overview of the MSAA Assessment System (see https://www.maine.gov/doe/sites/maine.gov.doe/files/inline-files/2016ParentOverview-allgradescombined.pdf).

Relevance: The evidence is highly relevant.

Completeness: Limited Evidence.

Overall Support: The evidence moderately supports the element.

Possible Challenges: Additional evidence needed; An example of additional evidence could be a survey of parents to begin to understand the degree to which parents use MSAA Science scores and other MSAA-based information to understand their child's progress from year to year.

11.3 Conclusions

Since 2023 is the second year of the MSAA Science assessment, many assumptions and associated elements have limited or moderate evidence. However, both the quantity and quality of the evidence are expected to improve in the following years.

The MSAA Psychometrics and Test Construction Subcommittee, which included committee members from the MSAA Science Partners, has developed and will implement a research agenda to develop evidence where it currently is missing or weak.

These assumptions and elements comprise the validity arguments for MSAA scores. Tables 11-3–11-5 summarize the rating scale for each assumption and element. The tables indicate the following:

Primary Score Intended Score Interpretation

MSAA scores provide reliable and valid information about important knowledge and skills in elementary, middle, and high school multidimensional science concepts that students with the most significant *cognitive disabilities are attaining*. Of the 21 assumptions and elements that support the intended score interpretation,

- 17 sets of evidence are highly relevant, 2 sets of evidence are moderately relevant, 0 sets of evidence are open to reasonable rebuttal or debate for the relevance rating scale.
- 8 sets of evidence are complete evidence, 7 sets of evidence are moderate to substantial evidence, 4 sets of evidence are limited evidence, and 0 sets of evidence are not relevant.
- 12 sets of evidence strongly support the claim or element, 2 sets of evidence moderately support the claim or element, 4 sets of evidence provide limited support to the claim or element, 0 sets of evidence do not support the claim or do not exist.

Intended Score Use 1

Schools and districts use the MSAA and its results to (a) monitor trends in student performance and (b) design professional development for teachers. Of the four assumptions and elements that support intended score use 1:

- 2 sets of evidence are highly relevant, and 2 sets of evidence are moderately relevant.
- 1 set of evidence is complete evidence, 1 set of evidence is moderate to substantial evidence, and 2 sets of evidence are limited evidence.
- 0 sets of evidence strongly support the claim or element, 4 sets of evidence moderately support the claim or element, and 0 sets of evidence provide limited support to the claim or element.

Intended Score Use 2

Teachers use the MSAA and its results to better integrate assessment with their instructional planning. Of the three assumptions and elements that support intended score use 2:

- 3 sets of evidence are moderately relevant.
- 3 sets of evidence are limited evidence.
- 3 sets of evidence moderately support the claim or element.

Intended Score Use 3

Parents use the MSAA and its results to get information about (a) what their child knows and what their child can do and (b) their child's progress from year to year. Of the four assumptions and elements that support intended score use 3:

- 1 set of evidence is highly relevant, and 3 sets of evidence are moderately relevant.
- 4 sets of evidence are limited evidence.
- 4 sets of evidence moderately support the claim or element.

Table 11-2. Status of Relevance of the Evidence for All Four SIUs, Assumptions, and Elements

Element	Highly Relevant	Moderately Relevant	The relevance of the evidence is open to reasonable rebuttal or debate
Primary Intended Score Interpretation MSAA scores provide reliable and valid information about important knowledge and skills in elementary, middle, and high school multidimensional science concepts that students with the most significant cognitive disabilities are attaining.			
1.1.1. The Extended Performance Expectations are aligned to the standards (aka, Performance Expectations) from A Framework for K-12 Science Education and, as such, aligned to each partner's academic content standards for each grade level.	X		
1.1.2. The 2023 MSAA Science items are aligned to the Extended Performance Expectations.	X		
1.1.3. The MSAA Science partners have confirmed that the Science Extended Performance Expectations, which are assessed on the MSAA Science, are aligned with each partner's academic content standards for each grade level.	X		
1.1.4. The 2023 operational MSAA Science items are aligned to the MSAA performance level descriptors.		X	
1.2.1. Items require application of the KSAs of the targeted construct.	X		

continued

Element	Highly Relevant	Moderately Relevant	The relevance of the evidence is open to reasonable rebuttal or debate
1.2.2. Items are accessible to all students.	X		
1.2.3. Appropriate accommodations are provided to meet student needs.	X		
1.2.4. Scaffolding is not a source of construct-irrelevant variance.	X		
1.2.5. Item rendering does not interfere with student access to test content.	X		
1.2.6. Platform does not interfere with student interaction with test content.	X		
1.2.7. Items are free of bias and sensitive issues.	X		
1.3.1. Test Administrators and School and District Coordinators understood and performed their roles properly.	X		
1.3.2. Test security concerns were limited.	X		
1.4.1. MSAA scores and categorizations into provisional performance levels are adequately reliable for their intended purpose.	X		
1.4.2. Item characteristics support intended interpretations about all students who take the MSAA.	X		
1.4.3. Test characteristics support intended interpretations about all students who take the MSAA.	X		
1.4.4. Scaling of the MSAA supports intended interpretations about all students who take the MSAA.	X		
1.4.5. Equating of MSAA test forms supports intended interpretations about MSAA students.			
1.5.1. Machine-scored items were scored accurately.	X		
1.6.1. MSAA scores correlate as expected with other measures of student proficiency.		X	
Primary Intended Score Use 1 Schools and districts use the MSAA and its results to (a) monitor trends in school performance and (b) design professional development for teachers how to monitor trends.			
2.1.1. MSAA Science scale scores for groups of students are adequately reliable and valid to enable school, district, and state leaders to monitor changes in means, standard deviations, and proficiency level percentages for classroom, school, district, and state groups.		X	
2.1.2. MSAA Science 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		
2.1.3. The relationship between MSAA Science scores and external measures of student achievement and growth is as expected, compared to grade-level assessments and other alternate assessments.	X		
2.2 (Assumption) MSAA Science results are used to design professional development for teachers.		X	
Primary Intended Score Use 2 The MSAA and its results are used to help teachers integrate MSAA scores and other information with their instructional planning.			
3.1.1. Teachers find the performance level descriptors and their students' performance levels useful for planning instruction, especially students in performance levels 1 and 2.		X	
3.1.2. Teachers find their students' scale score information useful for planning instruction, especially students in levels 1 and 2.		X	
3.2.1. Teachers use MSAA Science scores and other information for planning instruction.		X	
Primary Intended Score Use 3 Parents understand and interpret MSAA scores and other information correctly to understand what their child knows and can do.			

continued

Element	Highly Relevant	Moderately Relevant	The relevance of the evidence is open to reasonable rebuttal or debate
4.1.1. Parents understand and interpret correctly MSAA Science scores and other information to understand what their child knows and can do.		X	
4.1.2. Parents use MSAA Science scores and other information appropriately to understand what their child knows and can do and make decisions about their child's education and learning needs.		X	
4.2.1. Parents understand and interpret correctly MSAA Science scores and other information to understand their child's progress across grades.		X	
4.2.2. Parents use MSAA Science scores and other information appropriately to understand their child's progress across grades and make decisions about their child's education and learning needs.	X		

Table 11-3. Status of Completeness of the Evidence that Supports All Four SIUs, Assumptions, and Elements

Element	Completeness of the Evidence to the Element			
	Complete Evidence	Moderate to Substantial Evidence	Limited Evidence	No Evidence
Primary Intended Score Interpretation MSAA scores provide reliable and valid information about important knowledge and skills in elementary, middle, and high school multidimensional science concepts that students with the most significant cognitive disabilities are attaining.				
1.1.1. The Extended Performance Expectations are aligned to the standards (aka, Performance Expectations) from A Framework for K-12 Science Education and, as such, aligned to each partner's academic content standards for each grade level.	X			
1.1.2. The 2023 MSAA Science items are aligned to the Extended Performance Expectations.	X			
1.1.3. The MSAA Science partners have confirmed that the Science Extended Performance Expectations, which are assessed on the MSAA Science, are aligned with each partner's academic content standards for each grade level.	X			
1.1.4. The 2023 operational MSAA Science items are aligned to the MSAA performance level descriptors.			X	
1.2.1. Items require application of the KSAs of the targeted construct.	X			
1.2.2. Items are accessible to all students.		X		
1.2.3. Appropriate accommodations are provided to meet student needs.		X		
1.2.4. Scaffolding is not a source of construct-irrelevant variance.		X		
1.2.5. Item rendering does not interfere with student access to test content.		X		
1.2.6. Platform does not interfere with student interaction with test content.		X		
1.2.7. Items are free of bias and sensitive issues.		X		
1.3.1. Test Administrators and School and District Coordinators understood and performed their roles properly.			X	
1.3.2. Test security concerns were limited.			X	
1.4.1. MSAA scores and categorizations into provisional performance levels are adequately reliable for their intended purpose.	X			
1.4.2. Item characteristics support intended interpretations about all students who take the MSAA.	X			
1.4.3. Test characteristics support intended interpretations about all students who take the MSAA.		X		
1.4.4. Scaling of the MSAA supports intended interpretations about all students who take the MSAA.	X			
1.4.5. Equating of MSAA test forms supports intended interpretations about MSAA students.				

continued

Element	Completeness of the Evidence to the Element			
	Complete Evidence	Moderate to Substantial Evidence	Limited Evidence	No Evidence
1.5.1. Machine-scored items were scored accurately.	X			
1.6.1. MSAA scores correlate as expected with other measures of student proficiency.			X	
Primary Intended Score Use 1 Schools and districts use the MSAA and its results to (a) monitor trends in school performance and (b) design professional development for teachers how to monitor trends.				
2.1.1. MSAA Science scale scores for groups of students are adequately reliable and valid to enable school, district, and state leaders to monitor changes in means, standard deviations, and proficiency level percentages for classroom, school, district, and state groups.			X	
2.1.2. MSAA Science 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			
2.1.3. The relationship between MSAA Science scores and external measures of student achievement and growth is as expected, compared to grade-level assessments and other alternate assessments.		X		
2.2 (Assumption) MSAA Science results are used to design professional development for teachers.			X	
Primary Intended Score Use 2 The MSAA and its results are used to help teachers integrate MSAA scores and other information with their instructional planning.				
3.1.1. Teachers find the performance level descriptors and their students' performance levels useful for planning instruction, especially students in performance levels 1 and 2.			X	
3.1.2. Teachers find their students' scale score information useful for planning instruction, especially students in levels 1 and 2.			X	
3.2.1. Teachers use MSAA Science scores and other information for planning instruction.			X	
Primary Intended Score Use 3 Parents understand and interpret MSAA scores and other information correctly to understand what their child knows and can do.				
4.1.1. Parents understand and interpret correctly MSAA Science scores and other information to understand what their child knows and can do.			X	
4.1.2. Parents use MSAA Science scores and other information appropriately to understand what their child knows and can do and make decisions about their child's education and learning needs.			X	
4.2.1. Parents understand and interpret correctly MSAA Science scores and other information to understand their child's progress across grades.			X	
4.2.2. Parents use MSAA Science scores and other information appropriately to understand their child's progress across grades and make decisions about their child's education and learning needs.			X	

Table 11-4. Status of Overall Support for All Four SIUs, Assumptions, and Elements

Element	Overall Support to the Element			
	Strongly Support	Moderately Support	Limited Support	Does not Support
Primary Intended Score Interpretation MSAA scores provide reliable and valid information about important knowledge and skills in elementary, middle, and high school multidimensional science concepts that students with the most significant cognitive disabilities are attaining.				
1.1.1. The Extended Performance Expectations are aligned to the standards (aka, Performance Expectations) from A Framework for K-12 Science Education and, as such, aligned to each partner's academic content standards for each grade level.	X			
1.1.2. The 2023 MSAA Science items are aligned to the Extended Performance Expectations.	X			
1.1.3. The MSAA Science partners have confirmed that the Science Extended Performance Expectations, which are assessed on the MSAA Science, are aligned with each partner's academic content standards for each grade level.	X			
1.1.4. The 2023 operational MSAA Science items are aligned to the MSAA performance level descriptors.			X	
1.2.1. Items require application of the KSAs of the targeted construct.	X			
1.2.2. Items are accessible to all students.	X			
1.2.3. Appropriate accommodations are provided to meet student needs.	X			
1.2.4. Scaffolding is not a source of construct-irrelevant variance.	X			
1.2.5. Item rendering does not interfere with student access to test content.	X			
1.2.6. Platform does not interfere with student interaction with test content.	X			
1.2.7. Items are free of bias and sensitive issues.	X			
1.3.1. Test Administrators and School and District Coordinators understood and performed their roles properly.			X	
1.3.2. Test security concerns were limited.			X	
1.4.1. MSAA scores and categorizations into provisional performance levels are adequately reliable for their intended purpose.			X	
1.4.2. Item characteristics support intended interpretations about all students who take the MSAA.	X			
1.4.3. Test characteristics support intended interpretations about all students who take the MSAA.		X		
1.4.4. Scaling of the MSAA supports intended interpretations about all students who take the MSAA.		X		
1.4.5. Equating of MSAA test forms supports intended interpretations about MSAA students.				
1.5.1. Machine-scored items were scored accurately.	X			
1.6.1. MSAA scores correlate as expected with other measures of student proficiency.			X	
Primary Intended Score Use 1 Schools and districts use the MSAA and its results to (a) monitor trends in school performance and (b) design professional development for teachers how to monitor trends.				
2.1.1. MSAA Science scale scores for groups of students are adequately reliable and valid to enable school, district, and state leaders to monitor changes in means, standard deviations, and proficiency level percentages for classroom, school, district, and state groups.		X		
2.1.2. MSAA Science 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		
2.1.3. The relationship between MSAA Science scores and external measures of student achievement and growth is as expected, compared to grade-level assessments and other alternate assessments.		X		

continued

Element	Overall Support to the Element			
	Strongly Support	Moderately Support	Limited Support	Does not Support
2.2 (Assumption) MSAA Science results are used to design professional development for teachers.		X		
Primary Intended Score Use 2 The MSAA and its results are used to help teachers integrate MSAA scores and other information with their instructional planning.				
3.1.1. Teachers find the performance level descriptors and their students' performance levels useful for planning instruction, especially students in performance levels 1 and 2.		X		
3.1.2. Teachers find their students' scale score information useful for planning instruction, especially students in levels 1 and 2.		X		
3.2.1. Teachers use MSAA Science scores and other information for planning instruction.		X		
Primary Intended Score Use 3 Parents understand and interpret MSAA scores and other information correctly to understand what their child knows and can do.				
4.1.1. Parents understand and interpret correctly MSAA Science scores and other information to understand what their child knows and can do.		X		
4.1.2. Parents use MSAA Science scores and other information appropriately to understand what their child knows and can do and make decisions about their child's education and learning needs.		X		
4.2.1. Parents understand and interpret correctly MSAA Science scores and other information to understand their child's progress across grades.		X		
4.2.2. Parents use MSAA Science scores and other information appropriately to understand their child's progress across grades and make decisions about their child's education and learning needs.		X		