Multi-State Alternate Assessment

2023 ELA & Mathematics Technical Report

3/6/24

Prepared by Cognia for MSAA

TABLE OF CONTENTS

Appendix References by Chapter

Chapter	1	2	3	4	5	6	7	8	9	10	11	12
Appendix		A, B	C	D, E		F, G	G	H	I, J, K, L, M	M, N	O, P	

Chapter 1. Overview of MSAA and 2023 Updates

1.1 Purposes and Uses of the MSAA

The Multi-State Alternate Assessment (the MSAA) is a comprehensive, two-stage adaptive, 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 is designed to measure grade-level academic content that is aligned with, and derived from, MSAA Partner States' content standards. This test contains many built-in supports that 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 is administered in the areas of English language arts (ELA) and mathematics in grades 3–8 and High School (HS).

The MSAA Partners' long-term goal is to ensure that students with the most significant cognitive disabilities achieve increasingly higher academic outcomes and leave high school with the potential to pursue productive post-secondary options. A well-designed summative assessment alone is insufficient to achieve this goal. The MSAA is a component of a system of curriculum, instruction, and professional development that enables students with the most significant cognitive disabilities to access grade-level content aligned with grade-level state content standards.

The MSAA is 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 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 state content standards. To ensure that MSAA measures student achievement of alternate academic achievement standards aligned to grade-level content standards, this technical report provides the psychometric analyses and descriptions of technical procedures commonly found in all state assessment technical reports. In addition, this report identifies four primary intended interpretations and uses of MSAA scores and cites the assumptions and evidence that are relevant to those interpretations and uses.

1.2 Intended MSAA Score Interpretations and Uses

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

Primary Intended MSAA Score Interpretation

MSAA scores provide reliable and valid information about understanding the extent to which students with the most significant cognitive disabilities demonstrate important knowledge and skills in grade-level numeracy and literacy.

Primary Intended MSAA Score Uses

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

2. Teachers use the MSAA and its results to integrate assessment with their instructional planning.
3. Parents/Families 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.

The intended score interpretation and uses stated here align with the original statements of intended score interpretations and uses in the *National Center and State Collaborative 2015 Operational Assessment Technical Manual*. (See the "claim" and "uses" statements on page 8.)

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

1.3 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. Chapter 11 integrates this evidence into a validity argument for the intended MSAA test score interpretations and uses.

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 from the following five areas: test content, response processes, internal structure, relationship to other variables, and consequences of testing. These sources address different aspects of evidence to support 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.

1.4 Updates for the 2023 Program

The MSAA Partners for 2023 comprise American Samoa, Arizona, Bureau of Indian Education, Department of Defense Education Activity (DoDEA), Maine, Montana, Guam, The Commonwealth of the Northern Mariana Islands (CNMI), South Dakota, Tennessee, the U.S. Virgin Islands, Vermont (new 2023), and Washington, D.C.

New for 2023, MSAA items that followed the recommendations provided by an external consultant who collaborated with the MSAA Partners in 2019 were operational for the first time. Additional detailed

information about this process is available in Chapter 3. Also new for 2023, MSAA expanded the number and type of sample items available to teachers through the online assessment platform. In mathematics, selected-response or constructed-response sample items were added at each grade. In ELA, new passage sets (including a passage and related items), writing stand-alone items, and two open-response writing prompts (Level 2 and Level 3) at grade 6 were added. Each grade has its own Directions for Test Administration (DTA) that corresponds with the items in the online system to emulate and standardize the student testing experience.

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*. These documents and the online training modules were revised to streamline information and provide more clarity to Test Administrators (TAs) and Test Coordinators (TCs). Additional detailed information is available in Chapter 5.

Chapter 2. Overview of the MSAA

The MSAA assesses ELA and mathematics at grades 3–8 and HS and is aligned with the state content standards and the MSAA Core Content Connectors (CCCs). The MSAA is a computer-based, on-demand, two-stage adaptive assessment consisting primarily of selected-response items, along with some constructed-response items and open-response writing prompts. These item types are written at distinct levels of complexity, representing different levels of skill and knowledge acquisition by students.

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. As students learn and develop mastery of that skill or knowledge, they need less support.

The MSAA levels of complexity are designed to follow instructional practices. The test items are developed with many scaffolds and supports embedded within the items. Students are provided additional supports based on their individual requirements, including other allowable ways for Test Administrators (TAs) to present each item.

The MSAA 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, including assessment features built into the platform and accommodations such as using assistive technology, a scribe, and/or sign language. Appendix A contains the 2023 summary of accommodation usage frequencies for the MSAA. 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.

Mathematics consists of 35 operational items, primarily selected-response with some constructed-response items. ELA consists of 39–42 operational items, consisting primarily of selected-response items, some constructed-response items, a multiple-part selected-response writing prompt, and an open-response writing prompt at each grade level. Each content area assessment is stage adaptive with two stages, Session 1, and Session 2. The seven-week MSAA administration window allows for the test administrator to determine how many items a student completes in one sitting. There are also embedded field-test items in Session 1 for each grade and content area.

2.1 History of the MSAA

Work leading up to the MSAA began in late 2010, when the National Center and State Collaborative (NCSC) began development of the NCSC Alternate Assessment, which was designed to meet the requirements of the Elementary and Secondary Education Act and the Individuals with Disabilities Education Act and is based on alternate academic achievement standards for students with the most significant cognitive disabilities. This work culminated in the operationalized NCSC assessment in spring 2015. The work of NCSC ended following the spring 2015 administration. For additional information about the NCSC assessment, please refer to the *National Center and State Collaborative 2015 Operational Assessment Technical Manual* (see “References” for URL) or contact the MSAA Partners at MSAA@azed.gov. The MSAA Partners continued the work of NCSC, following many of the same principles, purposes, and core beliefs. The first administration of the MSAA was in the spring of 2016. Notably, the MSAA was not administered in 2020 due to school closings in response to the vast impact of COVID-19 on individuals worldwide.

2.1.1 Core Beliefs

The core beliefs that underlie the MSAA began with NCSC and were laid out in the prior planning and development of the AA-AAAS. As recorded in the *National Center and State Collaborative 2015 Operational Assessment Technical Manual*, states and organizational partners implementing the NCSC development plan found they needed to come to a consensus on topics that were a mix of practice and theory in the comprehensive context of teaching and learning. A blend of policy, educational, and technical solutions was required. Through policy discussions and in iterative research and design steps, the partners arrived at a shared philosophy and guiding principles that were reflected in the overall project resources. These project resources included a comprehensive system of curriculum, instruction, classroom assessment, and professional development as well as the operational assessment design.

The MSAA Partners believe, as their NCSC counterparts did before them, that accessibility is central to the validity argument of the assessment, and that access to content based on college- and career-ready academic standards begins with a rigorous curriculum, instruction resources, and training for teachers. The original design of NCSC curriculum and instruction resources was informed by extant research and iterative small studies to ensure inclusive accessibility and appropriately high expectations for learning. Then, the NCSC assessments were based on the same model of learning as reflected in classroom resources. Finally, the NCSC project provided 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 what they can do on the assessment. The NCSC Theory of Action and Validity Approach, available at ncscpartners.org/Media/Default/PDFs/Resources/NCSCBrief9.pdf, was developed to explain the basis for these resources, how they were intended to relate to one another and to college- and career-ready academic standards, and, ultimately, how they relate to the goal of having all students with the most significant cognitive disabilities leave high school ready to participate in college, careers, and their communities.

Practice-focused summaries of the foundational components reflected in the design of the NCSC assessment, known as the NCSC Brief series, are available to orient readers to the larger context of the comprehensive NCSC system of curriculum, instruction, assessment, and professional development. The NCSC Brief series can be found in the *National Center and State Collaborative 2015 Operational Assessment Technical Manual* (see “References” for URL) located here: http://www.ncscpartners.org/TechnicalDocumentation or by contacting the MSAA Partners at MSAA@azed.gov.

2.1.2 Stakeholders

Many stakeholders are involved in the development of the MSAA. MSAA Partners are key representatives who together compose the decision-making body for MSAA. Members of this body participate in various subcommittees that focus on specific aspects of the assessment and have decision-making authority on behalf of the MSAA Partners for each subcommittee’s focal area. Table 2-1 illustrates 2023 state representation for each subcommittee and is followed by a description of each subcommittee’s area of responsibility.

Table 2-1 Subcommittee Representation

Subcommittee	State Representation
Item Development	American Samoa, Arizona, the Bureau of Indian Education, CNMI Maine, South Dakota, Tennessee, and U.S. Virgin Islands
Manuals, User Guides, and Training	Arizona, Guam, Montana, South Dakota, Maine, and Tennessee
Psychometric and Test Construction	American Samoa, Arizona, the Bureau of Indian Education, CNMI, Maine, Montana, and South Dakota
Platform	Arizona, CNMI, Maine, Montana, Tennessee, and South Dakota
Scoring	Arizona, the Bureau of Indian Education, CNMI, Guam, Maine, Montana, and South Dakota
Reports	Arizona, the Bureau of Indian Education, Guam, Maine, and South Dakota

The MSAA Item Development Subcommittee provides overall input and direction related to development of field-test items; reviews all item development tasks; participates in development planning, and item and passage reviews; reviews alternative text; participates in Accessible Portable Item Protocol (APIP) reviews, plus the computer-based and paper-based materials review; and provides direction on updates to the graphics and editorial style guides, teacher directions, and front matter for the *Directions for Test Administration* (DTA). In addition to the Item Development Subcommittee, stakeholders from schools and districts across the MSAA Partners participate in the field-test item development process during the passage content and bias meeting and the item content and bias meeting. Additional detailed information is available in Chapter 4.

The Manuals, User Guides, and Training Subcommittee oversees 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*, online training modules, and the final quiz that is required for TAs.

The Psychometric and Test Construction Subcommittee oversees planning Technical Advisory Committee (TAC) meetings, contributes to psychometric decisions, reviews item performance statistics for each field-tested item during Data Review, approves the test design, approves the test construction test blueprint, provides decisions and approvals related to the MSAA constructed sets (operational and field-test items), provides the content of the End-of-Test Survey, determines relevant policies, receives the survey results after administration, and advises on the structure of the technical report.

The Platform Subcommittee determines development priorities for the online assessment platform each year and on an as-needed basis. This group also reviews recommendations and development pertaining to the security of the online platform and ultimately approves all changes made to the platform.

The Scoring Subcommittee reviews and approves the scoring specifications and scorer training materials, observes scoring processes, reviews daily scorer quality control and production management reports, and participates in daily debriefs during operational scoring.

Finally, the Reports Subcommittee is responsible for decisions pertaining to report revisions and design. This group also approves all changes made to the overall layout of the student results files and the final processing and reporting business requirements implemented for MSAA reporting. This subcommittee also reviews and approves the Parent Guides in English and Spanish and the *Guide for Score Report Interpretation*.

2.2 MSAA Participation

The criteria for student participation in the 2023 MSAA reflect the pervasive nature of a significant cognitive disability. All content areas are considered when determining who should participate in this assessment. Table 2-2 below shows the participation criteria and the descriptors used to determine eligibility for each student.

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 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.
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.

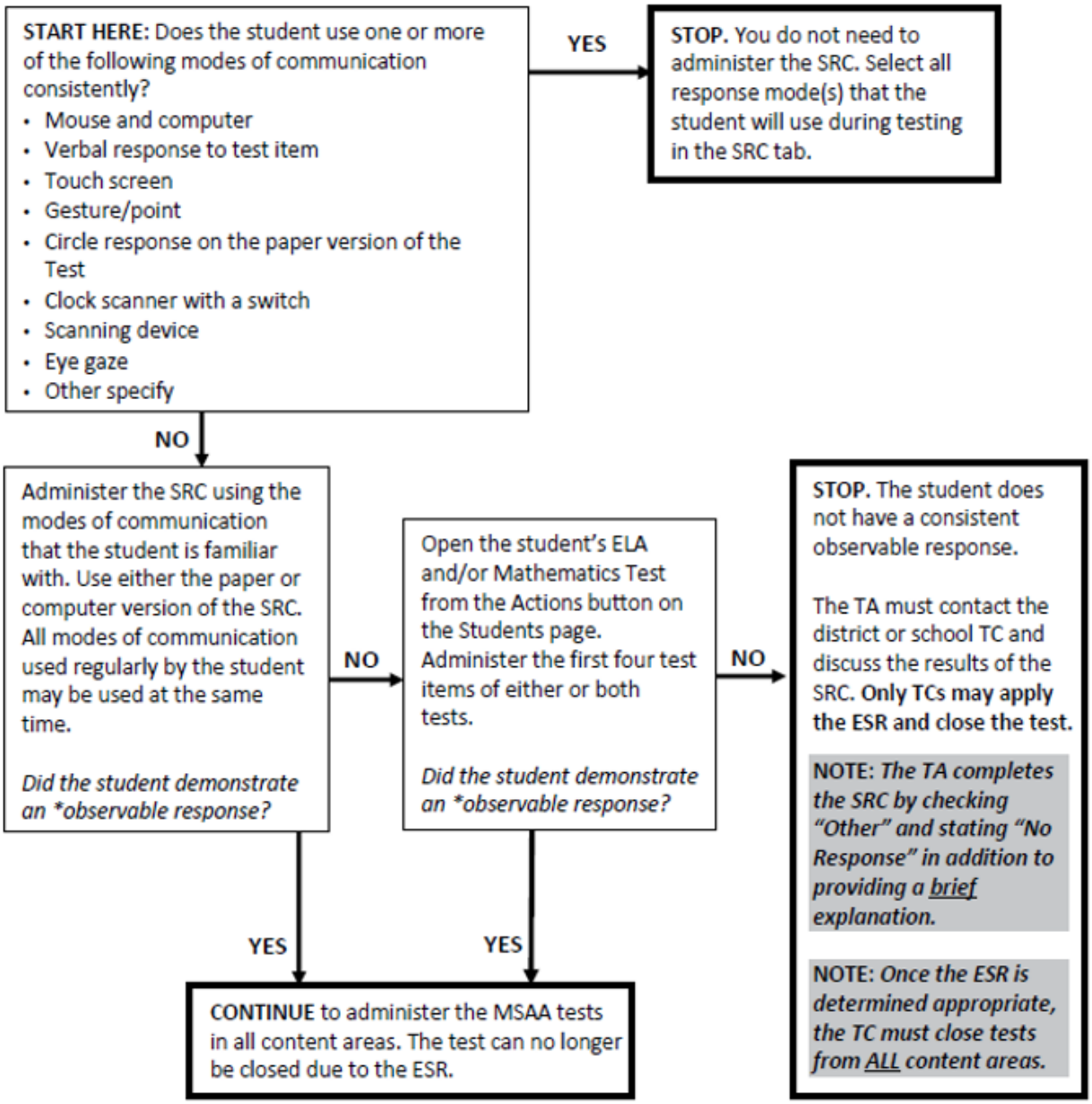
**Adaptive behavior is defined as essential for someone to live independently and to function safely in daily life.*

Appendix B shows the 2023 summary of students who participated in the MSAA for both mathematics and ELA by demographic category.

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 an 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.

Figure 2-1 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 that can be used throughout testing. If clear, intentional communication is not shown, the ESR may be applied. This figure shows the process of implementing the ESR.

Figure 2-1. Student Response Check (SRC) Flowchart: When to Apply the ESR



Post-assessment, teachers have the opportunity to use the Communication Tool Kit developed by NCSC to help these students develop an appropriate mode of communication. 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 Alternate Academic Achievement Standards and Core Content Connectors

As noted in Chapter 2, MSAA has evolved from the work of NCSC. As such, MSAA's history is firmly planted in the foundation of the NCSC AA-AAAS and follows the original Theory of Action. Designed specifically for students with the most significant cognitive disabilities, the NCSC AA-AAAS was a performance-based test that was aligned with grade-level state content standards for ELA and mathematics and tested student performance based on alternate academic achievement standards.

The NCSC state and center partners, comprised of content and special education experts, focused on defining the constructs of reading, writing, and mathematics to reflect an appropriate expectation of instruction and learning throughout a student's educational experience. Furthermore, the experts sought to make those constructs adaptable to the way in which students with the most significant cognitive disabilities demonstrate acquired knowledge and skills. NCSC established overarching content definitions by examining: (a) existing content definitions in general education; (b) the content, concepts, terminology, and tools of each domain; (c) a body of extant research; and (d) the Common Core State Standards (CCSS). These content definitions became central to the development of assessment items.

NCSC developers revised and refined the NCSC AA-AAAS design using cycles of continuous feedback from state and center partners. Developers evaluated proposed designs through iterative item and test development steps, special studies, and pilot testing, all of which were central to the final NCSC assessment model implemented through the first administration of the operational test in spring 2015.

Prior to the start of item development, the Core Content Connectors (CCCs) connecting the Learning Progression Frameworks (LPFs) to the CCSS were developed.

3.1.1 The Learning Progression Frameworks

The LPFs present a broad description of the essential content and general sequencing for student learning and skill development (Hess, 2010). The LPFs provide the educational logic to help move students with the most significant cognitive disabilities along with their peers, based on researched teaching and learning, toward mastering skills for college and career readiness. Experts at NCSC looked at these learning targets together with grade-level content expectations from the CCSS to identify and clarify the most salient grade-level core academic content to guide instruction and assessment from kindergarten through high school for students with the most significant cognitive disabilities. This core academic content is referred to as the CCCs.

3.1.2 Core Content Connectors

The CCCs were defined by NCSC as the academic content designed to frame the instruction and assessment of students with the most significant cognitive disabilities. This identified core content serves as a connection or stage between the LPFs (designed for typically developing students) and the CCSS (which define grade-level content and achievement). The CCCs are intentionally dually aligned with both.

The CCCs are designed to contribute to a fully aligned system of content, instruction, and assessment that focuses on the core content, knowledge, and skills needed at each grade to ensure success at the next grade level.

Each CCC represents a teachable and assessable part of the content. Related CCCs are addressed during instruction to create deeper understanding of grade-specific academic content. The CCCs are specifically intended to promote success as students advance alongside peers without disabilities from grade level to grade level. They are the starting point for instruction, they do not represent everything an individual student can and should learn.

The CCCs preserve the sequence of learning outlined in the LPFs, to the extent possible, while deconstructing the progress indicators (which describe concepts and skills along the learning continuum for each grade span in the learning progression) into smaller segments of content. The CCCs and corresponding *Curriculum Resource Guides* were developed to help explain and promote how students can engage in the CCSS while following the LPFs. To demonstrate the content sequence maintained by the CCCs, Table 3-1 shows a series of CCCs developed for multiple grades by NCSC for one big idea within the mathematics strand of geometry.

Table 3-1. CCCs Developed for Geometry Big Idea: Shapes and Figures—Their Attributes, Properties, and Corresponding Parts

Grades	Geometry Core Content Connectors			
K–2	K.G.M1a1 Recognize two-dimensional shapes (e.g., circle, square, triangle, rectangle) regardless of orientation or size	K.GM.1a2 Recognize two-dimensional shapes in environment regardless of orientation or size	K.GM.1a3 Use spatial language (e.g., above, below) to describe two-dimensional shapes	2.GM.1a4 Identify two-dimensional shapes such as rhombus, pentagon, hexagon, oval, and equilateral, isosceles, and scalene triangles
3–4	3.GM.1h1 Identify shared attributes of shapes	4.GM.1h2 Classify two-dimensional shapes based on attributes (number of angles)		
5–6	5.GM.1a1 Recognize properties of simple plane figures	5.GM.1b1 Distinguish plane figures by their properties		
7–8	7.GM.1e Construct or draw plane figures using properties	8.GM.1g1 Recognize congruent and similar figures		
HS	H.GM.1e Make formal geometric constructions with a variety of tools and methods	H.GM.1b Use definitions to determine congruency and similarity of figures		

The CCCs reference the *Learning Progressions Frameworks Designed for Use with the Common Core State Standards in Mathematics K–12* (Hess, 2010). The letter/number in each box provides a cross-reference to the letter/number in the original learning progressions. For example, for *3.GM.1h1*, the *3* means third grade, the *GM* means geometry, the *1h* relates to the specific progress indicator in the original learning progression, and the *1* means that it is the first in a series of connectors.

Table 3-1 shows how learner understanding builds across years. For example, in the second column, the student recognizes shapes, then compares shapes based on attributes, then distinguishes plane figures by properties, then recognizes congruent/similar figures, and finally by high school can use definitions to determine congruency/similarity of figures. These skills all promote the big idea about shapes—their attributes, properties, and corresponding parts (Wakeman, Lee, & Browder, 2012).

The MSAA Partners adopted the CCCs as the standards that the students are instructed and assessed against as a participating consortia member. These CCCs are the measured academic content on the MSAA and are aligned to and derived from each participating state's content standards.

3.2 Alignment and Linkages

Evidence that 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, NCSC conducted a series of studies to answer several key questions related to the alignment of the assessment. These efforts were meant to ensure that students' scores can be interpreted as reflecting the knowledge and skills defined in the standards and claims (developed by NCSC, see *National Center and State Collaborative 2015 Operational Assessment Technical Manual*). The alignment questions were:

1. What is the degree of alignment between the CCCs and the grade-level CCSS?
2. What is the degree of alignment between instructional student learning expectations and measurement targets (expectations for assessed knowledge and skills)?
3. To what degree do the assessment tasks and items align to the grade-level CCSS?
4. To what degree do the assessment tasks and items align to the performance-level descriptors (PLDs)?
5. How well do the claims align with grade-level content and provide useful information for tracking student progress toward achieving the knowledge and skills in the grade-level standards?

To address the five alignment questions, various studies were conducted between 2012 and 2015 at different points in the development process to ensure system coherence. Table 3-2 lists the studies, when each was conducted, and the alignment question being addressed. Each MSAA partner may have conducted their own reviews in addition to those listed below.

Table 3-2. Studies Related to Evidence of System Coherence

Study	Conducted	Claim for Which Evidence Is Provided
Relationship Studies	Mathematics: Summer 2012; Reading: Winter 2013; Writing: Summer 2013	Evidence for Alignment Question #1. The content and skills in the CCCs represent an adequate and appropriate sample of the grade-level CCSS.
UMASS Study of Coherence	Fall 2013	Evidence for Alignment Question #2. The targets for measurement provide information useful for tracking student progress in the CCSS and to teachers for providing instruction focused on academic expectations.
Task/Item Alignment Study	Summer 2015	Evidence for Alignment Question #3. The content and skills assessed by the NCSC AA-AAAS represent an adequate and appropriate sample of the grade-level CCSS.
Item Mapping Study	Summer 2015	Evidence for Alignment Question #4. The score reports are accurate and support appropriate inferences about student knowledge and skills.
Vertical Coherence Study	Summer 2015	Evidence for Alignment Question #5. The targets for measurement provide information useful for tracking student progress in the CCSS and for providing instruction focused on academic expectations.

MSAA has carefully and gradually evolved from NCSC, ensuring the alignment has been maintained while still allowing for adjustments. For example, the prioritized CCCs and reporting categories have remained the same for mathematics and for ELA (apart from one adjustment to the Reading Foundational CCC at grades 3 and 4) from those originally identified by NCSC. Section 3.4 provides details on the contents and blueprints. The various alignment studies noted in Table 3-2 are applicable for the MSAA, as the NCSC Theory of Action serves as the foundation for the MSAA program. MSAA has implemented test design adjustments that were outlined by NCSC during the original test design planning phase. One example of this is the implementation of the stage-adaptive test design. Section 3.3 below provides detailed information about the assessment design.

3.3 2023 MSAA Assessment Design

3.3.1 Operational Design

The operational MSAA is designed to produce valid and reliable mathematics and ELA scores for the intended uses. The mathematics and reading portions of the test are composed primarily of selected-response items. In mathematics, all grade levels also include constructed-response items that require students to work through a process to solve a problem. Writing is composed of selected-response stand-alone items, a multiple-part selected-response writing prompt, and an open-response writing prompt.

The operational items vary in complexity following the *Mathematics Tier Guidelines*, *ELA Tier Guidelines,* and *Passage and Item Sloping Guidelines,* where each tier at a given standard addresses both the content complexity and the degree of scaffolding and support provided with the items. The items assess grade-level academic concepts defined by either the focal knowledge, skill, and ability (FKSA; tiers 2-4) or Essential Understanding (EU; tier 1). The FKSAs allow for items of graduated complexity where the same FKSA is addressed but with increased levels of support and/or decreased levels of complexity so that students with various levels of cognitive ability can access the content. The EUs define entry-level skills based on a grade-specific CCC that builds increasing understanding of the grade-level content. The tiers provide four decreasingly complex versions (items) of the task referred to as Tier 4 (most complex), Tier 3 (less complex), Tier 2 (less complex than Tier 3), and Tier 1 (least complex). The writing prompts use

three levels of items. Level 1 is a multiple-part selected-response item series, where all items build on each other toward the creation of a final product. Level 2 and Level 3 are open-response writing prompts that vary in complexity based on the amount of support provided at each level.

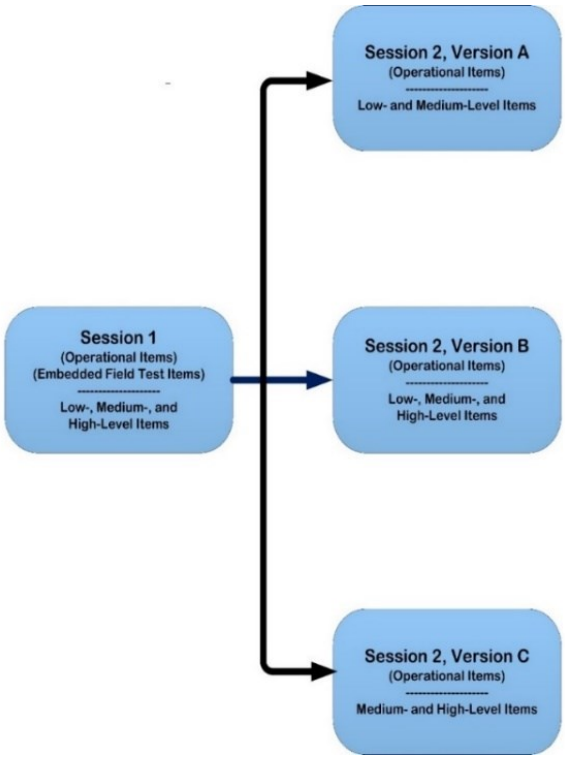
The MSAA TAC and Partners initially monitored item performance across 4 tiers but found they didn't sufficiently differentiate. They transitioned to 3 levels in item development. Tier 4 adhered to existing guidelines, Tier 1 remained mostly the same with minor adjustments, and Tiers 2 and 3 merged with changes to option presentation and modeling demonstrations aligned with FKSAs. This shift still aligned with the original complexity intent. Level 3 became the most complex, Level 2 less complex, and Level 1 the least complex. Writing prompt levels remained unchanged. Updated guidelines provided a structure for item design. More details can be found in Section 3.3.3.

For the 2023 assessment, three two-stage adaptive forms were developed for both ELA and mathematics to accommodate the inclusion of field-test items within Session 1. The forms follow guidelines informed by the respective content-area test blueprints (test blueprints are discussed in Section 3.4). The operational items are presented in Session 1 and Session 2. The Session 1 operational items are the same across all forms. Session 1 is considered Stage 1. Session 1 is taken by all students, while Session 2, which is considered Stage 2, is assigned to students based on how they perform on Session 1. There are three versions of Session 2, of varying difficulty, that may be assigned. Version C is intended to be slightly more complex and difficult than Version B, and Version B is intended to be slightly more complex and difficult than Version A. A Level 1 writing prompt is included for Sessions 2A, 2B, and 2C. A Level 2 writing prompt is included in Session 2A, and a Level 3 writing prompt is included in Sessions 2B and 2C. There are, thus, three possible paths for a student to take through the multistage test. All students take Stage 1, and, depending on how they perform on Stage 1, are assigned 2A, 2B, or 2C. In 2023 there is a moderate overlap of items in each version, but enough variation to ensure varying degrees of the desired separation of test information functions (TIFs) across the paths.

Students are routed to the appropriate path based on the routing cuts. For more information on TIF, see Chapter 9, and for more information on the measurement reliability stemming from these TIF values, see Chapter 10.

Figure 3-1 illustrates the two-stage adaptive design with field-test items indicated in Session 1, along with the levels of items that were used in each session. The three paths (Session 1 plus Session 2A, Session 1 plus Session 2B, and Session 1 plus Session 2C) for the operational assessment exist for each of the three field-test forms.

Figure 3.1 Two-Stage Adaptive Design



3.3.2 Operational Items and Embedded Field-Test Items

As discussed earlier, there are three versions of Session 2. Several items are the same across Sessions 2A, 2B, and 2C.

As shown in Table 3-3, the ELA tests administer 39–42 operational items including two writing prompts, across the testing sessions per grade. Additionally, each form has 11–13 field-test items for a total of 33–39 items across the three forms. As discussed earlier, there are three versions of Session 2, which consist of 3 passage sets and 5–7 writing items. In Session 2 a total of 7–21 items overlap across Sessions 2A, 2B, and 2C, depending on the grade. The items that overlap are not always the same ones across all 3 versions (e.g., one passage set may overlap across 2A and 2B, and a different passage set may overlap across 2B and 2C).

Table 3-3. ELA Operational and Embedded Field-Test Items

Grade	Total Operational Items Administered to Each Student	Writing Prompt Operational Items	Field-Test Items Total Across Three Field-Test Forms
3	42	2	30
4	42	2	36
5	40	2	32
6	39	2	34
7	39	2	32
8	39	2	36
HS	39	2	32

As shown in Table 3-4, the mathematics tests consist of 35 operational items across the testing sessions per grade. Additionally, each field-test form has 10 different field-test items for a total of 30 field-test items across the three field-test forms. As discussed earlier, Sessions 2A, 2B, and 2C each have 20 items. A subset of the items in Session 2A are common with items in Session 2B. There can be up to 10 items that are common between Sessions 2A and 2B. A subset of the items in Session 2C is common with items in Session 2B. There can be up to 10 items that are common between Sessions 2C and 2B. There are no common items between Session 2A and Session 2C.

Table 3-4. Mathematics Operational and Embedded Field-Test Items

Grade	Total Operational Items Administered to Each Student	Field-Test Items Total Across Three Field-Test Forms
3	35	30
4	35	30
5	35	30
6	35	30
7	35	30
8	35	30
HS	35	30

The 2023 field-test items were selected according to the following criteria:

- mathematics and ELA items represent a variety of item complexity levels (including the writing stand-alone component and a Level 1 writing prompt);
- ELA passage or writing topics are unique to the form and provide a variety of genres; and
- the passage and items are engaging, accurate, and free of regional bias.

The items on each of the forms are reviewed by psychometricians for any statistical concerns. The *AY23 Test Construction Process* provides 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. The content specialists create the test forms based on the test blueprints and criteria provided by the psychometricians. The forms are 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 Psychometric and Test Construction Subcommittee. All constructed tests, as well as the field-test items, are posted on a secure FTP site for the MSAA Psychometric and Test Construction Subcommittee review and approval. A webinar is held with the MSAA subcommittee to explain the test construction process and to review the Test Construction Design document, which provides information specific to each content area about the items selected. The MSAA subcommittee then has an opportunity to provide input and final approval.

The 2023 assessment includes field-test items in both mathematics and ELA (reading and writing) with differing levels of complexity. To address some of the overlap in item difficulty across Tiers 2, 3, and 4, the MSAA Partners adjusted the item design to collapse Tiers 2, 3, and 4, into Levels 2 and 3 (as described earlier and displayed in Table 3-5 below). Beginning in 2021 the items that were field-tested were written to three target levels. The operational items in the 2023 test forms include some items written to four target tiers and some items written to three target levels. The writing prompts were already

developed at only three tiers, and as such, were renamed to designate them as levels instead of tiers. The tiers-to-levels mapping is shown in Table 3-5.

Table 3-5. Mapping of Tiers and Complexity Levels

Tier 4	Item complexity level 3
Tier 3	Item complexity level 2
Tier 2	
Tier 1	Item complexity level 1

The *ELA Level Guidelines, Mathematics Level Guidelines,* and item specifications for each grade and content were updated to reflect the three item complexity levels. A primary distinction among items written at item complexity levels 1–3 is (a) their connection to content standards, and (b) the scaffolded supports provided at each level.

- Complexity level 3 items target the Core Content Connectors, with minimal supports provided during item administration.
- Complexity level 2 items target Core Content Connectors, with content supports (e.g., graphics, examples, definitions) provided during item administration.
- Complexity level 1 items target Essential Understandings, with content supports (e.g., graphics, simplified language) and item supports (e.g., two response options provided during item administration).

3.3.3 Item Design and Administration

The MSAA 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 Partners follow NCSC's intentional assessment development process to address the targeted grade-level academic content linked to evidence-based curricular and instructional materials.

The MSAA content development processes address levels of cognitive and language complexity, specifically addressing the state content standards, and the heterogeneous characteristics of the target student population. The assessment items vary systematically in complexity yet remain aligned with the FKSA or the EU behind the CCCs. 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 (e.g., reminders, examples, and models) are provided to focus the student on the task and elicit a response without guiding the student's response.

Overall Item Structure

A range of item levels is developed for each CCC, as described in Section 3.3.1. 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 assess grade-level academic concepts defined by either the FKSAs or the Essential Understandings (EUs). Items follow the level guidelines and item specifications. As outlined in the level guidelines, items of graduated complexity address the same FKSA but provide increased levels of support and/or decreased levels of complexity, and at the lowest tier address the EU that has the most decreased level of complexity, and, also, as part of the item, provide the greatest level of support. Additionally, the MSAA item specifications are consistent with design patterns and task template

guidelines that were originally developed by NCSC. The item types, as outlined in the MSAA item specifications, are selected-response, multiple-part selected-response, constructed-response, and open-response. Regardless of tier or item type, all items include scripted teacher directives.

Mathematics and ELA (reading, language, and stand-alone writing) selected-response items are multiple-choice items where a student selects a response from three options (two options at Level 1); the answer is worth 0 or 1 point. ELA multiple-part selected-response items are multiple-choice items that are clustered together and connected to a single CCC. For each item, the student selects a response from three options (two options at Level 1); the answer is worth 0 or 1 point. The overall cluster could, then, be worth more than 1 point. There are two- and three-part items. A typical example of a multiple-part selected-response item would be an initial item in the cluster that asks the student to identify the main idea and then a second item that asks for a supporting detail. In contrast, the mathematics MSAA item specifications and tier guidelines require CCCs with multiple components to be addressed with unique items. For example, a CCC asking a student to identify and solve an equation might be evaluated using one item that requires the student to identify the correct equation for a word problem and a separate item that requires solving an equation.

Mathematics and ELA constructed-response items require the student to interact in some way with a set of materials to provide a response. These items are scored as correct or incorrect by the TA following the directions provided in the *Directions for Test Administration* (DTA). For example, students might construct a graph, solve a problem, or complete a table, chart, or graphic organizer. These items are worth 0 or 1 point because the items ask the student to show whether a single concept is understood.

ELA reading foundational items focus on comprehension skills and are administered as selected-response items worth 0 or 1 point. Students are asked to read five words and select the most appropriate word to complete the sentence provided.

ELA writing prompt items require students to compose a permanent product about a specific topic, following the writing process. The Level 1 writing prompt is a multiple-part selected-response item where the items build on each other toward the creation of a final product. For each item in the series, the student selects a response from two options, with the answer worth 0 or 1 point. Items may have four to six parts, depending on the grade.

Unlike Level 1, the Level 2 and Level 3 writing prompts are open-response writing prompts that vary in complexity with the amount of support provided at each level. The Level 2 writing prompt provides a graphic organizer and a template with sentence starters that a student utilizes to create a product based on information he or she included in the graphic organizer. The Level 3 writing prompt provides a graphic organizer and a template that does not have sentence starters; the student completes his or her product within the template based on information he or she included in the graphic organizer. For Levels 2 and 3, the student response is evaluated against a grade- and level-specific rubric. Open-response writing prompt items were developed for Levels 2 and 3 only. As outlined in Chapter 1, the writing prompt items are operational in each grade for the 2023 MSAA. For reference, the specific writing rubrics are included as an appendix in the *MSAA 2023 Guide for Score Report Interpretation*.

New Item Approaches Operational in 2023

In 2019, MSAA Partners and the TAC were presented with a series of recommendations from Diane Browder, distinguished professor of special education emeritus with over 200 publications in the field of special education and alternate assessments. Dr. Browder advised the NCSC project as well and has

deep knowledge of the underlying philosophy and goals for the assessment program. Her review, discussion with the partners and TAC, and recommendations are based on new research-based understanding on how students with significant cognitive disabilities build competence in the academic domains. Based on this feedback and in collaboration with the Item Development Subcommittee, several new item approaches were developed.

First, these recommendations were reviewed to determine whether their implementation would affect development and test administration documentation, including the item specifications, the level guidelines, the style guide, the TAM, training modules, and MSAA instructional resources.

Then, content developers conducted an item bank analysis, identifying standards that, based on data, were the most challenging for students. These standards were targeted for implementation of Dr. Browder's recommendations to create new approaches for assessing the most challenging constructs.

Most of the incorporated suggestions required temporary, drafted updates to the item specifications, only to be made permanent if supported by data. After these updates were approved by the Item Development Subcommittee, item development began. Items incorporating suggested new approaches were developed in the fall of 2020 and spring of 2021 and were reviewed by item review committees in the summer of 2021. These items were then field tested in the 2022 administration.

The new item types and recommended approaches from Dr. Browder included an increased emphasis on the following:

- Constructed-response items, including constructed-response items for ELA (The constructed-response item type had previously only been used for mathematics.)
- graphic organizers for both ELA and mathematics
- scaffolding "rules" or remember statements
- prescriptive approach to context familiarity by level where in Level 1 passage and item context is kept close to home and school and Levels 2 and 3 extend to community and global contexts
- simplified language in the item stems, table headings, and teacher directions

Following data review of these 2022 field-test items, it was determined that the new approaches should be permanently included in the item specifications and the items should be included in the operational test blueprint. The items were operational on the 2023 assessment for the first time. Please refer to the quantitative analysis provided in Chapter 12 for the outcomes of these new item approaches, demonstrating improvements in the test information function at various cutoff scores. This is an essential component of the test construction process.

Administration

For every grade level, the ELA and mathematics tests contain two test Sessions. Test Session 1 contains operational and field-test items. Students are then routed to Session 2A, 2B, or 2C in accordance with their performance on Session 1. TAs begin with Session 1 of either the ELA test or the mathematics test. Descriptions of the test sessions are shown in Tables 3-6 and 3-7.

Table 3-6. ELA Test Sessions

Session 1: ELA	Session 2: ELA, Includes Writing Prompts
Literary and informational reading passages and associated selected-response and constructed-response reading items	Literary and informational reading passages and associated selected-response and constructed-response reading items
Selected-response and constructed response writing stand-alone items	One multiple-part selected-response writing prompt
Reading Foundational items (grades 3 and 4 only)	One open-response writing prompt
Field-test items	

Table 3-7. Mathematics Test Sessions

Session 1: Mathematics	Session 2: Mathematics
Selected-response mathematics items	Selected-response mathematics items
Constructed-response mathematics items*	Constructed-response mathematics items*
Field-test items	

**Constructed-response mathematics items are dichotomously scored.*

3.3.4 Item Components

3.3.4.1 Selected-Response: Reading, Writing (Stand-Alone Items and Multiple-Part Selected-Response Writing Prompt), Mathematics

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. Every item is presented in the following order:

- item stimulus (which may include a passage, passage part, picture, graphic, or other illustration);
- item question; and
- response options presented in vertical or horizontal formation depending on the size of the response options.

Students 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 System. If a student has the scribe accommodation, the scribe enters the student-selected response on behalf of the student.

3.3.4.2 Constructed-Response: Mathematics, ELA

The secure grade-, content-, and form-specific DTA contains the directions as well as the materials and manipulatives needed by the TA to assess the student on the constructed-response items. The TA prints out the materials and manipulatives with which the student will interact. Each item is presented to the student in a standardized, scripted sequence of steps, culminating in the TA scoring the student's performance using the required scoring rubrics. The scoring rubrics provide scoring standards that must be used in evaluating student responses. The constructed-response item is scored by the TA as correct or incorrect based on the scoring rubric for that item. The TA enters the student constructed-response score into the MSAA System.

3.3.4.3 Open-Response: Writing Prompt

All open-response writing prompt directions and stimulus materials, including the response template, are included in the secure grade-, content-, and form-specific DTA. TAs print or prepare any writing stimulus materials that they would need to use for the test. The open-response writing prompt is presented to the student by the TA in a standardized, scripted sequence of steps.

The student, or a scribe, records the response to the writing prompt either on the response template in the online MSAA System or on the paper response template included in the DTA. If the student uses a paper version of the response template, the TA

- uploads the response template, including any annotations, into the MSAA System, or
- transcribes or types (exactly) the student's writing response, including any annotations, into the MSAA System.

If the student's writing response includes inventive spelling, hard-to-read penmanship, or use of symbols, TAs are directed to annotate the response so that it can be understood by an external scorer. For more information about scoring, see Chapter 6.

3.4 Content and Blueprints

The test blueprints followed by MSAA are consistent with the original NCSC Theory of Action, the evidence-centered design undertaken to develop the summative assessment, and with best practices in educational measurement. Tables 3-8 and 3-9 show the broad targets developed to guide the item development process and to inform test construction. The tables provide general guidance for identifying areas of emphasis in the development of the mathematics and ELA tests. The test blueprints in Appendix C incorporate the overall content distributions used for the development of the operational tests. Each grade level/content area is represented by a table that first describes the domain (e.g., operations and algebraic thinking) or text type (e.g., reading informational text), weights by domain and ELA strands and text types, CCC, item types, and number of items. To continuously improve the assessment following each administration, the items' statistics for each test in each grade and content area are revisited to balance both the content requirements of the blueprints and the psychometric characteristics of the items for the subsequent administration. The core set of operational items on each two-stage adaptive test is established from this balanced approach.

3.4.1 English Language Arts

For the 2023 MSAA, the ELA items in reading and writing are aligned with prioritized CCCs, which are in turn connected to the CCSS and state content standards, as well as to the LPFs. The distribution of ELA items related to various text types (e.g., literary, informational, and argument) aligns to the text type emphasis in reading and writing outlined in the CCSS and state content standards.

For the 2023 MSAA, reading comprehension assessment items are presented as a single selected-response or multiple-part selected-response item as described in Section 3.3.3.

In grades 5–8 and HS, some prioritized content standards require evaluation of content across more than one passage. These skills are measured using paired passage sets. All paired passages are written in the informational text type. Tables in the test blueprints identify which CCCs require paired passages.

In grades 3 and 4, the reading foundational content category addresses the anchor standard of fluency. In 2023, the reading foundational items are used operationally.

The three CCCs prioritized for writing at each grade level consist of one CCC operationally assessed by a multiple-part selected-response writing prompt and an open-response writing prompt, and two CCCs operationally assessed by selected-response writing stand-alone items. The selected-response writing stand-alone items are designed to assess discrete basic writing skills. The multiple-part selected-response writing prompt and the open-response writing prompt are designed to measure a student's ability to generate a permanent product to represent organized ideas specific to a writing mode, supported with details or facts to develop those ideas or clarify meaning, and the use of standard English conventions (for the open-response writing prompt only).

Table 3-8. Guidelines for Distribution of ELA Content by Grade Level

ELA Content Category	Gr 3	Gr 4	Gr 5	Gr 6	Gr 7	Gr 8	HS
Reading Literary	24–32%	24–32%	25–33%	21–30%	17–26%	17–26%	17–26%
Reading Informational	18–26%	18–26%	25–33%	26–34%	32–36%	32–36%	32–36%
Reading Vocabulary and Foundational (G3 and G4)	12–16%	12–16%	6–10%	9–11%	6–9%	6–9%	6–9%
Writing	36–38%	32–38%	31–40%	36–40%	36–40%	36–40%	36–40%

3.4.2 Mathematics

Mathematics items are aligned with prioritized CCCs, which are in turn connected to the CCSS and state content standards, as well as to the LPFs. Mathematical knowledge across the CCCs is assessed through selected-response items and constructed-response items. The need for constructed-response items is determined by the FKSA associated with a given CCC.

Table 3-9. Guidelines for Distribution of Mathematics Content by Grade Level

Mathematics Content Category	Gr 3	Gr 4	Gr 5	Gr 6	Gr 7	Gr 8	HS
Operations and Algebraic Thinking	28–32%	28–32%	9–11%				
Number and Operations Base Ten	17–23%	9–11%	34–40%				
Number and Operations Fractions	17–23%	28–32%	17–23%				
Measurement and Data	17–23%	17–23%	17–23%				
Geometry	9–11%	9–11%	9–11%	9–11%	17–23%	28–32%	9–11%
Ratio and Proportions				28–32%	34–40%		
Expressions and Equations				17–23%	9–11%	17–23%	
The Number System				28–32%	17–23%	9–11%	
Statistics and Probability				9–11%	9–11%	17–23%	17–23%
Functions						17–23%	
Algebra and Functions							47–52%
Number and Quantity							17–23%

In some cases, the selected FKSAs are best addressed by separating the skill into two parts, creating two unique items to fully address a single content standard. Tables in Appendix C identify which CCCs require two items.

In addition, there are items identified as not allowing the use of calculators. These items tend to be related to computation, where the construct being assessed would be masked using a calculator.

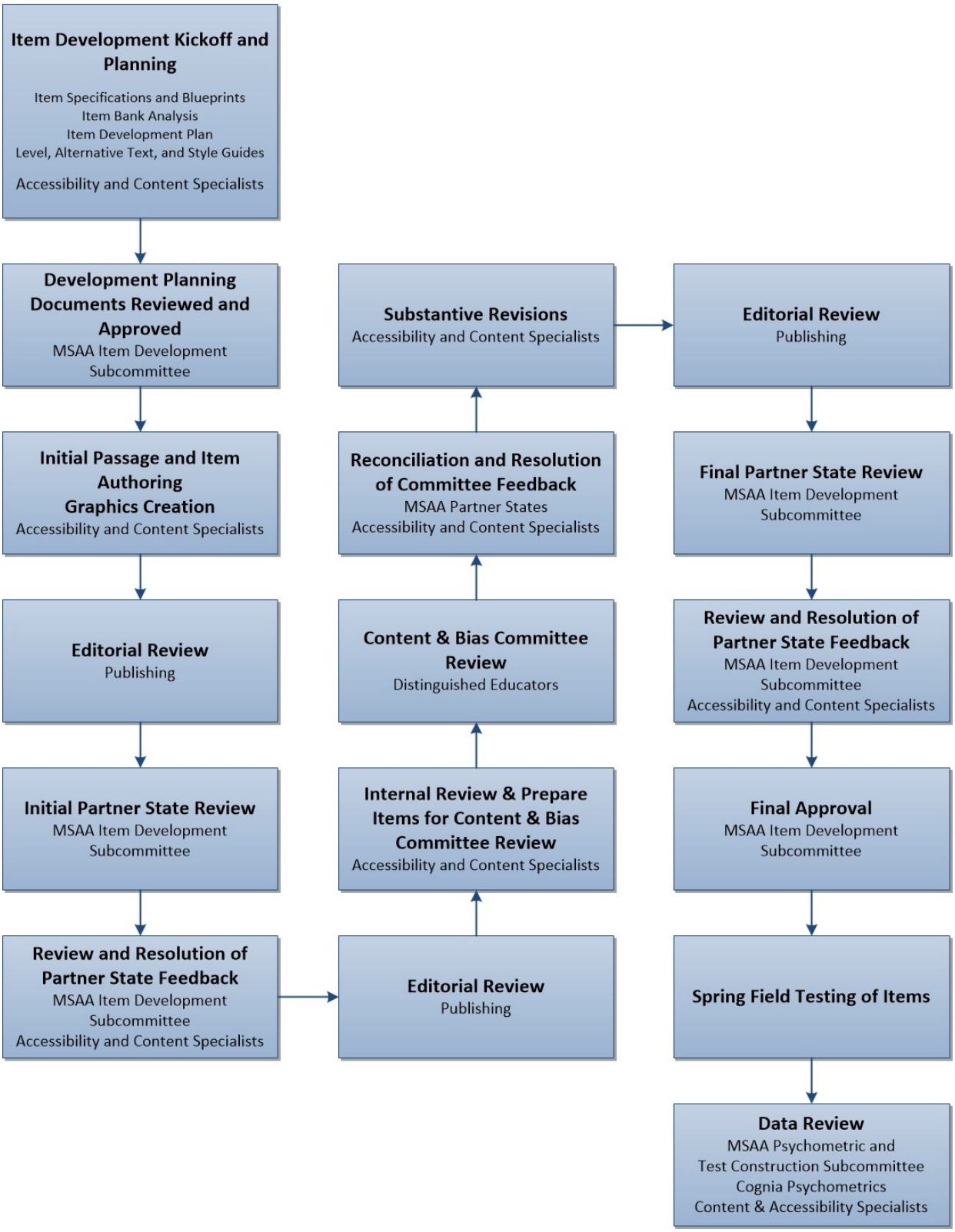
Chapter 4. Test Development-Stakeholder Involvement

4.1 General Philosophy and Role of the Item Development and Psychometric Subcommittees and Other Stakeholders in Test Development

As discussed previously, the MSAA is a comprehensive assessment system designed to promote increasingly 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 the academic content of the CCCs 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 two-stage adaptive assessment allows students to demonstrate what they know and what they can do. Given the wide diversity of the student population, great emphasis is placed on ensuring that the MSAA is appropriate and accessible to all eligible students.

The MSAA operational items on the 2023 administration are from the previous NCSC 2015 administration, as well as the 2016, 2017, 2018, 2019, 2021, and 2022 MSAA administrations. As described in Chapter 3, the items selected as field-test items are developed by MSAA. The item development process is an iterative one, which allows for multiple opportunities for review of the items by various stakeholders including MSAA Partners, content experts, and representative reviewers who are selected by MSAA Partners, and external passage and item content and bias review participants. 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 MSAA Partners. Figure 4-1 provides a flowchart outlining the item-development process.

Figure 4.1 Item Development Process



General and special education teachers, administrators, and other education specialists are selected to review passages for content or bias and sensitivity issues before item development begins for the ELA assessment. Additionally, an item content and bias review committee convenes in the summer to review newly developed items for content or bias and sensitivity issues in ELA and mathematics. Each ELA and mathematics content group reviews items for content-related considerations, such as alignment to the FKSA or EU, ratings of depth of knowledge, clarity of the item content, and consistency of teacher directions. Separate bias and sensitivity groups review the ELA and mathematics items for bias and sensitivity considerations, as well as accessibility considerations. The list of participants in the item content and bias review is included in Appendix D. Based on the review committee recommendations, 100% of the mathematics and ELA items were accepted or accepted with revisions by the participants in the item content and bias review meetings. All passages reviewed by the educator committee were accepted or accepted with revisions. Additionally, the passages and items were reviewed and approved by MSAA Partners.

The MSAA Item Development Subcommittee, which is made up of MSAA Partners, provides overall direction and guidance regarding field-test item development. This multistage development and review process by the subcommittee and educator review committees provides ample opportunity to evaluate items for their accessibility, appropriateness, and adherence to the principles of Universal Design. Documentation that guides these reviews has been developed and updated collaboratively with the Subcommittees throughout the life of MSAA. This documentation includes:

- item specifications
- level guidelines
- alternative text guidelines
- item review checklists
- bias checklists

Through these reviews, 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.

The MSAA Psychometric and Test Construction Subcommittee participates in the data review meeting(s) and is responsible for making determinations about the future usage of the items based on the field-test statistics. During the data review meeting(s) with the MSAA Psychometric and Test Construction Subcommittee, Cognia content specialists, accessibility specialists, and psychometricians review the Field-Test Calibration Report, which includes item statistics for each field-test item that has been flagged by the psychometricians. The statistics that trigger an item being flagged are shared with the subcommittee. Referenced during data review are the IRT analyses summarized in the Field-Test Calibration Report (see Section 9.2 for field-test calibration details). Data review attendees are also supplied with Asset Detail Reports, which provide the actual passage and item for each of the flagged items. This step allows for the content of the flagged items to be considered when determining future usage.

Flagged items are placed into categories. The items might be flagged only for form 2A (Use only in forms 2B and 2C), flagged for 2A and 2B (Use only in form 2C), flagged for all three 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 item is reviewed, the subcommittee members determine whether an item is accepted with the corresponding usage recommendation, rejected, or designated as revise and re-field-test. The following

tables provide a summary of the designations determined by the subcommittee based on the data from the AY22 field-test item performance.

Table 4-1. ELA AY22 Field-Test Data Review Summary

Grade	Accepted Usage Recommendation	Rejected (DNU)	Accepted Revise and Re-field-test
3	29	0	0
4	33	0	0
5	32	0	0
6	30	0	0
7	32	0	0
8	30	0	0
HS	32	0	0

Table 4-2. Mathematics AY22 Field-Test Data Review Summary

Grade	Accepted Usage Recommendation	Rejected (DNU)	Accepted Revise and Re-field-test
3	28	2	0
4	29	1	0
5	28	2	0
6	29	1	0
7	30	0	0
8	29	1	0
HS	29	1	0

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 (DNU) and designated as revise and re-field-test do not become part of the operational pool. It should be noted that this year presented the fewest number of items that have been rejected per grade that we have seen.

The MSAA Psychometric and Test Construction Subcommittee is also responsible for the review and approval of 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 Psychometric and Test Construction Subcommittee review and approval. A webinar is held with the MSAA subcommittee to explain the test construction process and to review the *AY23 Test Construction Process* document, which provides information specific to each content area about the items selected. The MSAA subcommittee then has an opportunity to provide input and final approval.

4.2 Sample Item Teacher Guide

A new resource was developed for use prior to the 2021 administration. Sample Item Teacher Guides were created to help teachers use the sample items as an additional assessment tool. This allowed teachers to understand what students may know and be able to do based on their performance on these sample items. They were able to respond to this information by applying instructional strategies and scaffolding suggestions outlined in the Teacher Guides. An example from the Sample Item Teacher Guides is included in Appendix E.

The Teacher Guides have a blueprint table at each grade that outlines the items in each sample test. The ELA blueprint table/overview can be used to help select the sample item(s) that will provide the best evidence of student learning. The learning targets differentiate between the types of evidence each item will provide. The item type describes how the student will engage with the item: through multiple choice,

constructed response, or open response. For ELA, items that address reading standards are grouped by passage set; each passage set primarily addresses standards in genre-specific content categories. The passages for items that assess reading standards are accessed in the *Directions for Test Administration* (DTA) and computer-based testing platform.

To obtain evidence of understanding for each grade-level standard, teachers can use the Teacher Guides to do the following:

- Access the sample items for the students' grade levels.
- Use items individually as the learning targets are covered in class.
- Use the items in small groups to address a series of learning targets that focus on one standard.
- Use the entire sample item set to measure students' understanding of learning targets before, during, or after instruction.
- Review sample item sets from lower grades to build understanding of prerequisite skills for a given standard.
- Review sample item sets from higher grades to know how standard and item information build from the target grade.
- Use the sample items as models to create additional items to assess the standards.

Chapter 5. Training and Administration

5.1 Test Administrator and Test Coordinator Training

The MSAA 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 the MSAA is a computer-administered test, the administration procedures are consistent with the hardware and software requirements of the test specifications. MSAA 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-, content-, and form-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. TCs and TAs receive both online training and supporting documents to ensure the fidelity of implementation and validity of the assessment result. Additionally, standardized training and supporting documents help MSAA Partners 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 are updated for the 2023 administration in conjunction with the updates to the required documents. There are six modules (see Table 5-1). Each module requires approximately 17-34 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

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. Each module must be viewed before the link for the subsequent module becomes accessible.

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. The TAs are notified within the MSAA System whether they pass the final quiz. They are allowed multiple attempts to obtain a score of 80% or higher on the final quiz. TAs are allowed access to the secure test materials only after fulfilling this certification requirement.

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 for the NCSC assessment, MSAA added sample items for the 2023 administration that are representative of current MSAA item development. The sample items do not address all assessed content at each grade level and are not representative of every item type. 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 throughout 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 six modules; each of which runs 17-34 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

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 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). Videos 3 and 4 focus on administration of the open-

response writing prompts. In each of these videos a mock student-TA interaction is used to provide TAs with a true picture of these administration processes (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
Video 3: How to Administer a Level 2 Writing Prompt
Video 4: How to Administer a Level 3 Writing Prompt

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, who 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;
- 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 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. Each DTA is accessible through the MSAA System once a TA has been certified. The DTAs must be used by the TA for MSAA administration. The following elements are provided as part of each DTA (as applicable for a content area):

- standardized directions and scripts to be followed exactly as written for each item, including alternative text as appropriate;
- details about manipulatives required to administer a test item, such as calculators and counters;
- reference sheets that contain important graphics;
- materials required for administration of the mathematics and ELA constructed-response items;
- scoring rubrics for mathematics and ELA constructed-response items;
- writing prompt scripts, graphic organizers, student response templates, and stimulus materials for all writing prompts in each grade-level ELA DTA; and
- specific directions on how to administer the braille versions of ELA foundational reading items in grades 3 and 4.

While the TA has some flexibility in presentation and response mode to ensure the MSAA 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 to navigate the MSAA System. Each user guide contains specific information relevant to the role of the TA and the TC. The user guides provide many efficient 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 was March 14–April 29, 2023. Both the ELA and mathematics assessments were completed within the same administration window. 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 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, 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 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, as part of the MSAA. Support is provided to TCs and TAs through the MSAA Service Center, additional supports built into the MSAA System functionality, and the MSAA Partner States. TA feedback is gathered through an End-of-Test Survey. Review of the service center logs, and analysis of the test survey results informs MSAA Partner States 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 8:00 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 MSAA Service Center ticketing system, ServiceNow, and is included in weekly status reports that are provided to MSAA Partners. These reports summarize ticket activity, call analysis

data (e.g., call duration, hold time), and per-grade/content and per-state 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 the field. Cases with policy or consortium-wide implications are directed to MSAA Partners and related policy documentation.

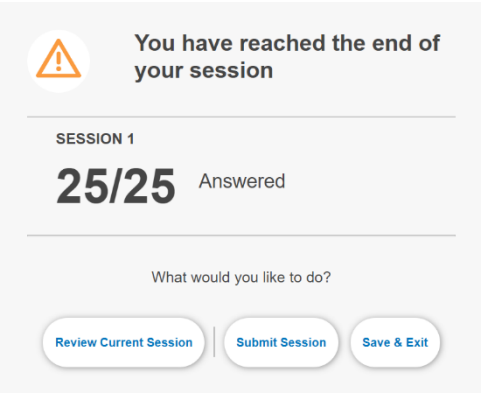
Furthermore, a banner messaging system in the MSAA System is implemented, as needed, to notify users of important information during the administration window. When the messaging system is activated, a banner message appears at the top of the screen upon login to notify users of system information and 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 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 professional codes of ethics and state 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 a session, a prompt appears notifying the TA that he or she has reached the end of the session, displaying the number of answered items, and presenting options for the TA to proceed to the next phase of the test (either Session 2 or final submission of the completed test, as appropriate), return to the current session, or save and exit the test.

Figure 5.1 End of Session 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 Session* button, the session is submitted and cannot be re-opened, and the TA is permitted to continue to the next phase of the test. This prompt reduces the risk of users accidentally submitting a session without properly understanding the implications.

Throughout the administration window, Cognia monitors activity and provides weekly updates to MSAA Partners on test status 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 creation of assessment-wide or state-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.

5.8.4 Operational Test Survey Results

An End-of-Test Survey (EOTS) allows MSAA Partners 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 survey questions focus on several themes:

- technology use in the classroom,
- student behaviors and engagement,
- instructional time spent on academic content, and
- available professional development.

The results of the EOTS highlight several areas of concern that the MSAA Partner States 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 strategies.

The survey data also identifies the effectiveness of several improvements implemented in the 2023 MSAA to correct issues identified in previous administrations. The results of this survey indicate that many teachers emphasize the writing process. Teachers shared that they include writing introductions and conclusions, using graphic organizers, drafting, editing, and revising in their instruction. However, most teachers also value practice with letter formation and copying words or sentences from a model. The MSAA Partners continue to train on the meaning of writing for students with the most significant cognitive disabilities and on using grade-level standards when teaching writing skills.

One issue raised by the teachers in the EOTS data is a lack of continuity between instruction and assessment. The MSAA Partner States 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 standardized administration.

Several questions on the survey address teachers' instructional practices for 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.

The results of this survey indicate that many teachers emphasize the writing process. Teachers shared that they include writing introductions and conclusions, using graphic organizers, drafting, editing, and revising in their instruction. However, most teachers also value practice with letter formation and copying words or sentences from a model. The MSAA Partners continue to train on the meaning of writing for students with the most significant cognitive disabilities and on using grade-level standards when teaching writing skills.

Furthermore, responses from TAs 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 incorporate test administration strategies for students with unique needs.

The EOTS data also show that many students are using a variety of Assistive Technology (AT) 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. Students taking the MSAA are using devices that promote communication and independence. The MSAA Partners independently use these results to make professional development decisions within their entities.

Chapter 6. Scoring

6.1 Selected-response and Constructed-Response Item Scoring Processes

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

Overview of Scoring Process Within the Assessment System

The MSAA System provides automated machine scoring for all item types, aside from the open-response writing prompt and mathematics constructed-response items, which require human scoring. The selected-response and constructed-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 teachers to enter responses for a student into the paper-based test delivery. The selected-response items are scored according to the answer keys provided in each test package. The constructed-response items are scored as a correct or incorrect student response, which is then entered by the Test Administrator (TA). At the completion of the operational test, all test data are extracted from the system and are then compiled to generate full result sets for each student's tests.

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 and constructed-response item keys were entered correctly. A key verification check is conducted by the data analyst. Any items that may be flagged are provided to the content specialists to conduct a blind key check. The content specialists review the actual item and mark the key in the flagged file. Any mismatches are researched by the content specialist, and updates are made following a problem-item notice 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 F.

Test Administrator/Scorer 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. When TA scoring is required, such as in the case of the mathematics constructed-response items, the DTA includes the teacher scripting and directions related to any item setup and administration specifics, any templates required by the items, and the rubrics used to score the items. Once the item is administered, the TA enters the response as correct or incorrect in the MSAA System.

The *MSAA Online Assessment System User Guide for Test Administrators* provides further directions to TAs on entering item responses in the MSAA. 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, scoring, entry of student responses, and submission of student responses during the administration window, TAs can call or e-mail the MSAA Service Center with any questions.

6.2 Open-Response Writing Prompts Scoring Processes

6.2.1 Overview of Open-Response Writing Entry Process Within the Assessment System and Test Administrator Training

Open-Response Writing Entry Process

As described in Chapter 1, the open-response writing prompts in grades 3–8 and HS are being operationally administered in the 2023 MSAA. The open-response writing prompts are described in detail in Chapter 3. The student, or a qualified scribe, records the response on either the response template in the MSAA System or the paper response template included in the writing DTA. TAs can upload the student's final writing response template directly in the system, retype the student response within the item response field of the item, or upload the template and retype it within the item response field. The item responses (no matter how they are entered) are then extracted from the online system and provided to Cognia for human scoring.

Test Administrator Training and Support

All TAs are required to participate in administration training modules and pass a final quiz to be certified to administer the MSAA assessment, as described in Chapter 5. The TA training includes review of the parameters for the administration of the open-response writing prompt, as well as entry of the student responses into the MSAA System. In addition, the best practice videos provide a student-TA representation that gives TAs a true picture of the processes involved in conducting the open-response writing prompt. During the test administration, TAs use the grade-, content-, and form-specific DTAs to administer each open-response writing prompt. The DTAs include the teacher scripting and directions related to any item setup, administration specifics, and the materials for the open-response writing prompt.

The *MSAA System User Guide for Test Administrators* provides further direction on entry of student responses to the open-response writing prompt. Additionally, the MSAA Service Center provides support for TAs.

6.2.2 Benchmarking and Identification of Scoring Materials

The open-response writing prompts were benchmarked during the 2015, 2016, and 2017 field tests. During the benchmarking activity, Cognia scoring experts (Scoring Supervisors and Scoring Team Leaders [STLs], defined below) worked collaboratively with NCSC representatives in 2015 and with MSAA representatives from the Scoring Subcommittee in 2016 and 2017 to review student responses, assign a score based on the MSAA grade- and level-specific rubrics for each trait (i.e., organization, idea development, conventions), and identify item-specific writing anchors and practice sets.

The final scores for the anchor and practice sets were recorded, and representatives from NCSC (2015) and the MSAA Scoring Subcommittee (2016 and 2017) acknowledged their consensus on the sign-off

document for each prompt. Also, development of a scoring decisions document began in 2017. It was reviewed by the MSAA Scoring Subcommittee, which provided rationale and decision points to be used during scoring by the Scoring Supervisors and STLs.

Following the identification of the anchor sets, two qualification sets were identified for each prompt. Each qualification set consisted of 10 responses; scores were based on anchor responses and scoring decisions made during the benchmarking meetings. The MSAA Scoring Subcommittee reviewed and approved the scores and responses used for qualification sets.

6.2.3 Scorer Recruitment and Qualifications

The MSAA scorers are a diverse group of individuals with a broad range of backgrounds, including teachers, business professionals, graduate students, and retired educators. They are primarily obtained through Cognia's Human Resources. Other temporary employment agencies may supplement the efforts of Cognia's HR group. All selected scorers hold the required minimum of a four-year college degree that includes ELA or writing coursework. 41% of the scoring team hold a master's degree and 8% hold a Ph.D. 50% of the scoring team assigned to the MSAA have previous scoring experience, and preference is given to those who have some experience with an alternate-assessment student population. All scorers sign a nondisclosure/confidentiality agreement.

6.2.4 Cognia Staff and Scoring Leadership

The MSAA operational open-response writing prompts were scored between April 24 and May 15, 2023. All Level 2 scoring occurred on-site at the Alpharetta Office. All Level 3 scoring was remote during office hours instead of at a regional scoring center.

The following staff members participated:

- Director, Scoring Operations: Primarily responsible for coordinating scheduling, budgeting, and logistics of all Scoring Centers. In addition, the Director for Scoring Operations coordinates the scoring of special education contracts, has overall responsibility for MSAA scoring-related activities, and serves as the Scoring Services Project Manager for MSAA.
- ELA Group Manager for Scoring: Responsible for managing scoring-related activities and monitoring reports, as well as leadership and training of scorers to ensure overall consistency of scoring.
- Scoring Content Specialist: Responsible for overseeing scoring activities across grades and monitoring accuracy and productivity across groups.
- Accessibility Assessment Specialist: Responsible for overseeing scoring activities and acting as the accessibility lead in coordination with the Cognia scoring staff.
- *iScore* Operations Manager: Responsible for setup and maintenance of *iScore* scoring system and for coordinating technical communication.
- Scoring Supervisor: Responsible for selecting calibration responses, training STLs and scorers, resolving arbitrations, and monitoring the consistency of scoring for items in assigned grades. Scoring Supervisors may also participate in benchmarking and identifying qualification sets prior to the onset of scoring.
- Scoring Team Leader (STL): Responsible for performing quality-control measures, resolving arbitrations, and monitoring the accuracy of a small group of scorers, usually consisting of not more than six. STLs may also participate in benchmarking and identifying qualification sets prior to the onset of scoring.

6.2.5 Training

Scoring Content Specialists and Scoring Supervisors assigned to train the STLs and scorers thoroughly review the decisions and materials that result from the benchmarking meetings in preparation for training.

One Scoring Supervisor is assigned to all Level 2 writing prompts across grades. Two Scoring Supervisors are assigned to the Level 3 writing prompts with a focus on lower or upper grades, The Scoring Content Specialists and Scoring Supervisors are responsible for creating prerecorded training modules for use in training. Leadership training for 2023 took place from April 17th through April 21st. STLs are required to meet or exceed the accuracy standard of 80% exact agreement on all items and at least 90% exact/adjacent[1] agreement on each trait. This requirement is applied to each of the three writing traits[2] individually across qualification sets 1 and 2. The STLs are also present during scorer training, which further reinforces their understanding of the rubrics and training materials.

Scoring Content Specialists and Scoring Supervisors conduct training on each open-response writing prompt before scorers are allowed access to student responses. Scorers are divided into three groups. One group focuses on Level 2 items and the other two groups on Level 3 items. Training sessions for scorers are facilitated by the Scoring Content Specialists and a Scoring Supervisor and are conducted in the following manner:

- Training commences with an introduction to scoring and an overview to explain the purpose and goal of the testing program and any unique features of the test and/or testing population.
- A general discussion addresses the security, confidentiality, and proprietary nature of testing, scoring materials, and procedures.
- Initial item training consists of a pre-recorded module that focuses on the following:
    - the three traits of the MSAA analytic rubrics for writing and how the scoring for each trait is applied to student work (See "Writing Scoring Rubrics," an appendix to the *MSAA 2023 Guide for Score Report Interpretation*, provided in Appendix G of this report.)
    - pertinent information on the testing instructions and item stimuli
    - actual responses with an item-specific anchor set, averaging 10 responses representing a range of scores across traits
    - anchor exemplars (presented in a predetermined order) that consist of responses that are typical, rather than unusual or uncommon; solid, rather than controversial or borderline; and true
    - the anchor response score and the scoring rationale, allowing scorers to internalize typical characteristics of each score point
- Scorers are instructed to refer to the anchor set frequently during scoring.
- After completing the module, training continues with the Scoring Content Specialist and/or the Scoring Supervisor presenting the supplementary training materials practice responses representing all score points across traits, when possible, and often containing responses that are more unusual and/or less solid (e.g., are shorter than normal, employ atypical approaches, or contain both very low and very high attributes). None of the practice papers contain responses that would require identification as nonscorable responses.
- During the review of practice responses, the trainer(s) often focus on the distinction between adjacent score points or clarification of other scoring issues that are traditionally difficult for scorers to internalize.
- After scorers independently read and score each practice response, the trainer(s) discusses the actual score and explains the rationale.
- A question-and-answer segment addresses any remaining questions from scorers and provides clarification prior to the qualification process.

[1] *"Adjacent agreement" means that the two scores differed by only one score point.*
[2] *The three writing traits are organization, idea development, and conventions. See rubrics embedded in Appendix G.*

6.2.6 Qualification

Following the training for each prompt, scorers are required to complete a qualification set to determine eligibility to score student work. There are two qualification sets in each grade and level consisting of 10 responses each. The responses, which represent a range of score points, are randomly distributed to scorers through *iScore*.

Scorers have two opportunities to qualify. If scorers attain a score match of at least 80% exact and 90% exact/adjacent agreement on all traits for the first qualification set, they are considered a "qualified scorer" and permitted to score live student responses. If they do not attain the required percentages, the Scoring Supervisor conducts retraining. Following this retraining, scorers are assigned qualification set 2. Since scorers qualify at the trait level, a scorer who qualifies on the first and third trait in qualification set 1, for example, receives the retraining referenced above. However, this scorer would only be required to qualify on trait 2 in qualification set 2. When the data indicates that a qualified scorer has demonstrated a weakness in a particular trait, that qualified scorer receives additional training prior to the start of scoring.

Scorers who fail to achieve the minimum levels of agreement are not allowed to score. When scorers demonstrate a level of understanding and the ability to apply feedback during the training and qualification process on a certain writing prompt, Scoring Leadership may choose to include the scorer in future trainings on a different writing prompt.

When the queue for the first open-response writing prompt is cleared, the training process is repeated for the next grade and level. This process continues until all 14 open-response writing prompts are scored. (See Section 6.2.1.) Scorer qualification success rates tend to improve as they train additional items/grades. They increase their understanding of the rubric and its use over time. In most instances, the initial team of scorers who began on the project remain on the project. However, some attrition occurs over the course of scoring and additional scorers are added to maintain the scoring schedule. Often, the Level 2 scorers will join the Level 3 scorers near the project's end and only after they have completed scoring of Level 2. The qualification results are listed in Table 6-1. This table includes the total number of scorers and leadership who attempted to qualify on the item and the total number who successfully did so. It also includes the percentage of people who passed all three traits on either the first or second qualification set combined.

Table 6-1. 2023 Scorer Qualification Rates

Grade	Level 2			Level 3		
	# Attempted	# Passed	% Passed	# Attempted	# Passed	% Passed
3	9	4	44%	17	6	35%
4	8	4	71%	17	13	76%
5	8	6	75%	14	12	86%
6	7	7	100%	13	11	84%
7	9	6	66%	17	7	41%
8	7	6	85%	12	5	42%
HS	7	7	100%	16	8	50%

6.2.7 Method for Scoring Operational Open-Response Writing Prompts

Student responses to the open-response writing prompts and any uploaded material are exported from the platform and imported to the Cognia *iScore* scoring system. Through *iScore*, qualified scorers read

and evaluate student responses, submitting scores electronically. The processes by which images are logged in, scanned, and uploaded into *iScore* provide anonymity to individual students and ensure random distribution of all responses during scoring.

All student responses are scored from uploaded evidence and/or computer-generated text, defined as student work directly entered in the MSAA System. For Level 2 prompts, when both uploaded and computer-generated text is available, the uploaded evidence is scored first, and the computer-generated text is used for clarification and confirmation of the uploaded student writing evidence. When there is only uploaded writing evidence but no computer-generated text to provide clarification and confirmation, the uploaded writing evidence is scored. When there is only computer-generated text but no uploaded writing evidence, the computer-generated text is scored. For Level 3 prompts, the computer-generated text and the uploaded evidence serve to provide a holistic demonstration of student ability and are considered together when both are available. When only one portion is available, the prompt is scored like a Level 2 prompt.

The following processes are in place during the scoring of the MSAA operational open-response writing prompts:

- The *iScore* system forces scorers to review all available pages before allowing a score to be submitted.
- All scoring is "blind." Only booklet numbers within *iScore* are linked to student responses; no student names are visible to scorers unless a name appears on material uploaded by the TA.
- Cognia maintains security during scoring by using a highly secure server-to-server interface to ensure that access to all student response images is limited to those who are scoring or working for Cognia in a scoring management capacity.
- During scoring, *iScore* enables constant measuring and monitoring of scorers for scoring accuracy and consistency. Each scorer's reading rate and total number of responses scored are also monitored.
- Scorers are required to maintain an acceptable scoring accuracy rate (80% exact and 90% exact/adjacent agreement) daily as measured through read-behinds, double-blinds, and daily calibration sets. (These measures are described below.)
- Scorers who repeatedly fall below standard are retrained or dismissed from scoring that item.
- Scoring rules are in place to determine the final score of record, or when a final score is to be provided by Scoring Leadership. (For examples of scoring resolutions, see Section 6.2.8.4.)

Table 6-2 represents the total number of student responses scored by writing prompt in each grade.

Table 6-2. Student Responses per Grade

	Number of Student Responses		
Grade	WRCC002	WRCC003	Total
3	1,069	1,141	2,210
4	1,083	1,040	2,123
5	1,066	1,252	2,318
6	745	1,541	2,286
7	721	1,572	2,293
8	1,014	1,344	2,358
HS	725	1,481	2,206

Note: For identification purposes in iScore, Level 2 prompts were designated as WRCC002 across all grades, and Level 3 prompts were designated WRCC003.

Scoring Rules

All open-response writing prompts are scored against a three-trait rubric (see rubrics in Appendix G). The scoring scale options of 0, 1, 2, and 3 are applied to each trait. (Note: for determining a student's total raw score to be transformed to a scaled score, the score categories of 1 and 2 were combined to be converted to a 1, and score category 3 was converted to a 2. These converted trait scores were the scores used in the psychometric analyses.) When a response does not conform to score point parameters, scorers can designate the response as one of the following:

- **Blank:** There is no attempt to respond to the item; no uploaded material is provided, and no response has been typed.
- **Unreadable:** The text on the scorer's computer screen is indecipherable or too faint to read accurately.
- **Non-English:** The response is written in a language other than English.
- **Repeats the Prompt:** The response is a direct copy of the prompt without any original text.
- **No Score:** The response requires clarification or adjudication by Scoring Leadership; scorers can assign this designation only with approval from Scoring Leadership.

Table 6-3 displays the resolution process for each of the responses described above.

Table 6-3. Scoring Resolution Process

Designation	Resolution Process
Blank	Responses scored Blank are sent to another scorer for a second read. Responses scored Blank twice are converted to zeros ("0") for reporting purposes. Any discrepancies are resolved by the Scoring Leadership.
Unreadable	Responses judged unreadable are forwarded to a special queue within *iScore* to be reviewed by a Scoring Supervisor, who resolves the student score. (If the response remains unreadable after review, the Scoring Supervisor assigns a score of "0.")
Non-English	Responses written in a language other than English are marked non-English and are converted to zeros ("0") for reporting purposes.
No Score	Responses that require additional clarification or adjudication are escalated to Scoring Leadership for response appraisal and scoring. This designation includes responses where more than one student's work appears to have been uploaded to the response. Responses where the uploaded evidence is a mismatch to the typed response are escalated to Scoring Leadership for response appraisal and scoring. Responses that legitimately respond to another item are escalated for review by Scoring Leadership. Any student response indicating administrative inconsistencies, potential cheating, and/or security lapses before, during, or after the test administration is scored based on its merits and then forwarded for review. If further attention is warranted, the State Services team notifies the appropriate MSAA Partner State. Responses determined to be nonscorable are resolved by the Cognia leadership team in consultation with the MSAA Scoring Subcommittee, if necessary.

Scorers also have the option of flagging a response as an "Alert," requiring immediate review and possible immediate action by Scoring Leadership and an MSAA Partner State. "Alert" responses can include, but are not limited to, those that suggest one or more of the following problems:

- thoughts of suicide
- criminal activity
- alcohol or drug use
- extreme depression
- violence
- rape, sexual or physical abuse
- self-harm or intent to harm others
- neglect

Scoring flagged ten responses as “Alerts” requiring immediate review during the scoring process. As a new procedure in 2023, Scoring also flagged 46 responses as an “Alert” due to the student image appearing in the uploaded material. All flagged responses were forwarded to the appropriate Partner State representative. See Table 6-4.

Table 6-4. Responses Flagged With “Alert”

MSAA Partner State	Number of “Alert” Responses Flagged as Possible Crisis	Number of “Alert” Responses Flagged for Student Image in Upload
Arizona (AZ)	1	16
District of Columbia (DC)	0	1
DoDEA (DD)	0	2
Guam (GU)	0	10
Montana (MT)	3	0
South Dakota (SD)	0	1
Tennessee (TN)	5	16
Vermont (VT)	1	0

6.2.8 Monitoring of Scoring Quality Control

Scorers are continuously monitored to ensure that scoring is accurate and consistent. Throughout the scoring process, read-behind scoring, double-blind scoring, and calibration sets are used as quality-control measures. MSAA Scoring Subcommittee representatives, along with the Cognia Accessibility and Scoring teams, monitor reports daily. Read-behind and double-blind statistics are reviewed daily. Calibration sets are administered and reviewed repeatedly during the project. Scoring Leadership and Content Specialists from the Scoring Services and Content Development/Accessibility departments at Cognia pay close attention to the disaggregated read-behind, double-blind, and calibration statistics.

Scorers needing more clarification on applying scores to specific traits are coached by Scoring Leadership. This continuous training allows Scoring Leadership an opportunity to resolve issues, reiterate scoring guidelines, and establish parameters for atypical student responses. Scorers who demonstrate inaccurate or inconsistent scoring are retrained and allowed to resume scoring under increased supervision. Scoring Leadership removes scorers who continue to fall below accuracy standards. On any day that a scorer falls below accuracy standards, the work is voided and rescored by other qualified scorers. During MSAA scoring the void process occurred 38 times.

6.2.8.1 Calibration Sets

To determine whether scorers are still calibrating to the scoring standard, they are required to complete a trio of online calibration sets at the start of each day, beginning with the second day of scoring. Scoring Leadership selects the responses for the sets, with each calibration set consisting of five responses

representing a range of scores. Scorers who assign at least 12 out of 15 scores exactly can then begin scoring for the day. Scorers who fail to meet that standard are retrained by discussing the calibration responses in terms of the rubric and the anchor set. Scoring Leadership determines if these retrained scorers should be allowed to begin scoring; though if they are, these scorers continue to be closely monitored. Over the course of scoring, 22% of all scorers (across all seven grades and 14 items), required retraining at least once. Scorers who received retraining successfully were permitted to resume scoring,

6.2.8.2 Read-Behind Scoring

Read-behinds provide a crucial tool in verifying scorer accuracy. STLs complete read-behinds on individual scorers on a daily basis. The STL's evaluation of each response is performed with no knowledge of the scores assigned across traits. The scores are only available to the STLs after they have also scored the response. If there is a difference in scores, either adjacent (one score point difference) or discrepant (more than one score point difference), the STL score is the score of record. If the scores are discrepant, or if there are a significant number of adjacent scores between the scorer and the STL, the STL discusses the rationale with the scorer.

The average number of read-behinds for each scorer is 5–10 reads a day, but this number varies depending on the accuracy of each scorer. 6.3% of all responses in this administration had a read-behind performed on the score. The read-behinds provide an immediate means of identifying scorers in need of further clarification on how to effectively apply the scoring rubrics to student responses. If scorers fall consistently below the 80% exact and 90% exact/adjacent (combined) threshold Scoring Leadership voids their scores for the day and may release them from scoring that item. Scoring Leadership monitors scoring accuracy and consistency by reviewing the read-behinds performed by the STLs and completing read-behinds on the STLs when possible.

6.2.8.3 Double-Blind Scoring

While read-behinds measure scorer accuracy in relationship to STL scores, double-blind scoring provides statistics on scorer-to-scorer agreement. Double-blind scoring is the practice of having two scorers independently score a response, without knowing either the identity of the other scorer or the score the other scorer assigned. In double-blind scoring, neither scorer knows which response will be (or already has been) scored by another randomly selected scorer. All responses for MSAA are 100% double-blind scored.

In addition to monitoring interrater agreement rates, double-blind scoring allows Scoring Leadership to resolve arbitrations when two scorers' double-blind scores do not agree across any of the three traits. If there is not exact agreement, *iScore* automatically places the response into an arbitration queue. Scoring Leadership, with no prior knowledge of the scores assigned, evaluates the response, with the leadership score becoming the score of record. The double-blind statistics provide an overview of agreement rate among the entire pool of scorers and assists in identifying any need for retraining.

6.2.8.4 Final Score Resolution

If scorers are adjacent in their scoring of a response, the two scores are averaged and rounded up for the score of record. If the scorers are discrepant in their scoring, the response will be sent to an STL for arbitration. The STL will review the response, provide the final score of record, and counsel scorers as needed. During the arbitration by scoring leadership, all three traits are evaluated and the final score of record for each trait is supplied by scoring leadership.

In read-behind cases, the Scoring Supervisor/STL score is the final score of record. For adjacent and discrepant scorer scores, the read-behind score is the final score of record. If a response gets more than

one read-behind and the two scores supplied by the STLs do not agree, a resolution score is needed. In the unlikely event that a resolution is required, the Scoring Supervisor provides a final score for all three traits during the post-scoring edit process.

6.2.9 Quality and Production Management Reports

Reports generated through *iScore* are essential during the scoring of the MSAA. Reports provide real-time statistics for review by the Cognia Scoring team and the MSAA Scoring Subcommittee to closely monitor scoring, thereby ensuring that

- scorer data (individual level) is monitored in real time to allow early scorer intervention when necessary;
- overall accuracy, consistency, and reliability of scoring (group level) is maintained;
- individual traits in need of further clarification are identified; and
- scoring schedules are upheld.

The reports listed in Table 6-5 provide the comprehensive tools and statistical information needed to execute quality control and manage production.

Table 6-5. Scoring Quality Control and Production Management

Report	Description
Read-Behind Disaggregated Summary	The Read-Behind Disaggregated Summary report shows the total number of read-behind responses read by both the scorer and the STL, and notes the number and percentage of exact, adjacent, and discrepant scores across each trait.
Double-Blind Disaggregated Summary	The Double-Blind Disaggregated Summary report shows the total number of double-blind responses read by a scorer and notes the number and percentage of exact, adjacent, and discrepant scores across each trait.
Compilation Report	The Compilation Report shows, for each scorer, the total number of responses scored, the number of calibration responses scored, and the percentage of exact, adjacent, and discrepant scores across each trait.
Summary Report	The Summary Report lists the total number of student responses loaded into *iScore*. This report includes the number of reads completed to date and the number of reads that remain.

6.2.10 Interrater Agreement

Kappa statistics (kappa coefficients) measure the agreement among two or more raters. The calculation is based on the difference between the level of agreement observed compared to the level of agreement that would be expected by chance alone. Kappa is a measure of this difference standardized to lie on a -1 to 1 scale, where 1 is perfect agreement, 0 is exactly what would be expected by chance, and negative values indicate disagreement. The kappa information in Table 6-6 shows agreement between raters at Substantial Agreement or Almost Perfect Agreement ranges for most of the open-response writing prompts across grades. In seven cases, the kappa agreement rate is in the Moderate Agreement range.

Table 6-6. Kappa Agreement—Operational Open-Response Writing

Grade	Item	Organization Trait 1	Idea Development Trait 2	Conventions Trait 3
3	WRCC002	0.64	0.64	0.84
	WRCC003	0.73	0.69	0.80
4	WRCC002	0.59	0.67	0.84
	WRCC003	0.79	0.75	0.83
5	WRCC002	0.60	0.61	0.82
	WRCC003	0.73	0.75	0.79
6	WRCC002	0.62	0.68	0.81
	WRCC003	0.64	0.59	0.82
7	WRCC002	0.61	0.60	0.70
	WRCC003	0.58	0.51	0.81
8	WRCC002	0.65	0.64	0.80
	WRCC003	0.61	0.61	0.77
HS	WRCC002	0.74	0.66	0.80
	WRCC003	0.61	0.54	0.77

Note: For identification purposes in iScore*, Level 2 prompts are designated as WRCC002 across all grades and Level 3 prompts are designated as WRCC003.*

Agreement Ranges:
< 0 Disagreement
0 = Chance Agreement
0.01–0.20 Slight Agreement
0.21–0.40 Fair Agreement
0.41–0.60 Moderate Agreement
0.61–0.80 Substantial Agreement
0.81–0.99 Almost Perfect Agreement

Chapter 7. Reporting

7.1 Development and Approval of Report Specific Documents

The *Reporting Services Deliverables Decision Rules* document ensures that reported results for MSAA 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 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.

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 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 Partner States, and subcommittee review and integration of feedback, until final revisions were approved by all participating MSAA Partner States.

Creating the Report Design Templates

To develop the report design templates, Cognia worked with the MSAA Reports Subcommittee to identify modifications to the templates used last year that would ensure that the data elements, layout, and report text were meaningful for reporting the spring 2023 MSAA results. Once finalized, the results of this collaborative process were presented to participating MSAA Partners for final approval. Changes were:

- Vermont became a Partner State and tested its 9th graders in ELA and Mathematics, while all other Partners tested their 11th graders in those two areas. Partner States agreed to change references to “Grade 11” in the reports to “High School.” (See 12.5 for more details on standards comparison)
- The “What to Work on Next” text in the Student Report was simplified.
- Print-Ready Student Report PDFs were no longer offered to Partner States to opt into. These are state-level PDF files that contain the Student Reports that Partner States’ constituents received.

MSAA 2023 Guide for Score Report Interpretation

Cognia uses an iterative process to annually update the *Guide for Score Report Interpretation* 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. States are permitted to remove codes not used in their state. Appendices included in this guide contain the Performance-Level Descriptors (PLDs) for ELA and mathematics, a sample individual student

report, and the writing prompt scoring rubrics. The final, approved *MSAA 2023 Guide for Score Report Interpretation* is delivered electronically to the MSAA 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, annually reviews and updates the following primary reports:

- 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 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 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 July to schools, districts, and parents as soon as possible at the beginning of the school year.

The primary results reported are the student's scale score and performance-level classification for mathematics and ELA. The performance-level classifications, with cut scores determined through the original standard setting and subsequent standards validation processes (see Chapter 9 for more information), are reported under the generic 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 comparing individual student performance to overall state performance and of school and district results with the overall state results.

BIE and VT both have state-specific options for additional, dynamic reporting services. BIE is contracted with eMetric and uses Lighthouse's Reporting/Data Interaction, which is eMetric's K-12 assessment

reporting and analytics platform. The disaggregate statistics displayed in the platform were provided to eMetric by Cognia. Vermont is contracted with FocalPoint for a similar service, using FocalPoint's reporting platform, LENS. Disaggregate statistics displayed by FocalPoint were provided to FocalPoint by Cognia. The Partner states all have access to the same disaggregate statistics in their state student results file, but only BIE and VT have additional platforms that allow them dynamic reporting functionality.

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 in at least one content area, as defined by the student report requirements. The report contains results for both ELA and mathematics content areas and was developed for parents and guardians of students who participated in MSAA. 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* and appear in this document in Appendix G.

Page 1 of the report contains the scale score, performance level, and associated performance-level descriptor for the level obtained by the student for each content area. A sentence below the graphical display explains the standard error of measurement (SEM) in layperson's terms by providing the expected range of scores the student would likely earn if tested again. For example, "If your child were to be tested again, it is likely that they would receive a score between 1228 and 1236."

Page 2 contains a brief overview of MSAA, 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.

Tests for students unable to show an 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.

If a student receives a student report but does not receive results for one of the two content areas, results for the missing content area are replaced with text encouraging parents or guardians to contact the child's teacher or school for more information.

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 MSAA, with a snapshot of their participation/test status and results for both content areas. The number of tested students, the average scale score, and the percentage of students by 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, MSAA Partners do not suppress information when the number of students participating is small. Additionally, these rosters are confidential to authorized school and district personnel only. 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 identified by name and state student identification number. It is intended that these data points be used in conjunction with the *MSAA 2023 Guide for Score Report Interpretation*. For each content area, 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 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 content area 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

Proprietary 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 the data are handed off to different teams within the IT-Reporting Department, the sending team verifies that the data is 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 is captured and delivered accurately. Quality control checks are being performed by the data processors and data analysts as the data is 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:

- Testing data counts of data imported.
- Testing data quality of individual fields for valid values, such as Gender, Ethnicity, etc.
- 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 is 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 have been conducted to ensure that the 2023 MSAA ELA (reading and writing) and mathematics 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 two parts: (1) differential item functioning (DIF) statistics and (2) dimensionality analysis of inter-item correlations. The item analyses presented here are based on the administration of the MSAA in spring 2023.

Note that classical difficulty (p-value) and discrimination (point-biserial) indices are often used in testing programs to compare the quality of items. However, such indices are not appropriate for a multistage adaptive test. The inappropriateness of these statistics stems from the fact that when two items are on two different stages or on different levels of a stage, the students taking one of the items will tend to have a higher overall ability distribution (as measured by scaled score) compared to the students taking the other item. As an example of the inappropriateness, consider a case where the two items have similar p-values, but one item is from Stage 2A, and the other is from 2C. This similarity would lead to the misleading inference that the two items are comparable in difficulty when, in fact, the 2C item is likely to be much harder than the 2A item. Thus, the classical difficulty and discrimination statistics are not included in the evaluation of item quality presented in this chapter. Chapter 9 provides IRT item parameter details for evaluating item difficulties.

8.1 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 items were evaluated in terms of DIF statistics.

For the 2023 administration, 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., cause 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, 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 H 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 MSAA items fell within this range (see Appendix H). Dorans and Holland further state 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.

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 Tables 8-1 and 8-2. Since an item can exhibit DIF for multiple comparisons, an item was counted once if any of the comparisons showed “high” DIF. Tables 8-1 and 8-2 show that only a few items were classified as “high” DIF for each grade and each level. These operational items continue to be monitored for content bias as well as evaluated for potential construct irrelevant factors and any issues to be addressed. These results indicate that the content bias reviews for ELA and Math were conducted thoroughly. Please note that items listed in Tables 8-1 and 8-2 are operational items, which means they have passed the initial content and bias review during the field test. These high DIF operational items were monitored closely by both Cognia content and psychometric teams once identified. For instance, if an item exhibits significant Differential Item Functioning (DIF) in year 1 but doesn't display such a pattern in year 2, that item will be retained in the item bank. Nevertheless, these items were not removed from the test form until further substantiating evidence was obtained to justify their removal.

Table 8-1. Number of Items with “High” DIF by Level—ELA

Grade	N_Item	Level 1	Level 2	Level 3
3	76	3	0	0
4	79	0	1	1
5	78	1	1	0
6	70	1	1	0
8	70	1	0	0

Table 8-2. Number of Items with "High" DIF by Level—Mathematics

Grade	N_Item	Level 1	Level 2	Level 3
3	69	0	2	0
4	67	0	2	0
5	70	0	1	1
6	68	0	1	0
7	71	0	1	0
11	68	1	0	0

8.2 Dimensionality Analysis

Because tests are constructed with multiple content-area subcategories, and their associated knowledge and skills, the potential exists for a large number of 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 2023 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 and if it was found: (a) estimated the size of the LID and (b) studied the nature of the LID with particular emphasis on possible multidimensionality. Our findings 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 displays the greatest evidence of LID. The cross-validation sample is then used to test whether the conditional covariances of the selected cluster of items display 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 (the size of the violation of LI). As with DIMTEST, the data are first divided into training and cross-validation samples. The training sample is used to find a set of mutually exclusive and collectively exhaustive clusters of items that best fit a 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 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. In 2018 an additional culling procedure was introduced to determine which data would be used to calibrate the operational items. The new procedure was introduced in response to the repeated finding in dimensionality analyses from previous years 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 same data were also chosen for the dimensionality analyses. Table 8-3 summarizes the dimensionality analysis sample sizes both prior to and after removing the R9-stringers. Table 8-3 also includes the percentages of stingers from the previous year as a point of reference.

Table 8-3. Summary of 2023 Testing Population

Subject	Grade	Total Before Removing Stringers	Total After Removing Stringers	Number of Stringers	Percent Stringers	Percent Stringers in 2022
ELA	3	2,190	1,853	337	15	16
	4	2,316	1,926	390	17	14
	5	2,305	1,933	372	16	14
	6	2,263	1,907	356	16	14
	7	2,266	1,944	322	14	14
	8	2,349	1,956	393	17	15
	HS	2,188	1,862	326	15	11
Mathematics	3	2,182	1,891	291	13	11
	4	2,308	1,993	315	14	10
	5	2,310	1,961	349	15	15
	6	2,263	2,009	254	11	9
	7	2,275	2,045	230	10	11
	8	2,345	2,085	260	11	10
	HS	2,180	1,942	238	11	8

DIMTEST and DETECT were separately applied to the three operational paths of each grade on the 2023 MSAA ELA and mathematics tests. The three paths resulted in three datasets to be analyzed for each ELA and mathematics grade-level test, a total of 42 analyses. First, each dataset was split into a training sample and a cross-validation sample. The sample sizes across the 42 analyses varied from a low of 457 (grade 4, mathematics, Path A) to a high of 911 (grade 4 mathematics, Path B). A rough tabulation of the sample size distribution is given in Table 8-4, including a comparison to the most recent past

administration in 2022. The sample sizes in 2023 were very similar to those in 2022. The overall average sample size for 2023 was about 650, whereas the overall average sample size for 2022 was about 635.

Table 8-4. Dataset Sample Sizes Used for Dimensionality Analyses*

Sample Size	Number of Datasets					
	Path A		Path B		Path C	
	2022	2023	2022	2023	2022	2023
< 600	7	10	4	3	4	2
600 to 800	6	4	8	8	9	9
800 to 1000	1	0	2	3	1	3
> 1000	0	0	0	0	0	0

**Stringers not included in these analyses.*

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

Table 8-5. Average Multidimensional Effect Sizes by Content Area and Grade Across Years*

Path	Content Area	Grade	Multidimensionality Effect Size	
			2022	2023
A	ELA	3	0.50	0.43
		4	0.47	0.36
		5	0.77	0.51
		6	0.58	0.28
		7	0.74	0.46
		8	0.47	0.46
		HS	0.33	0.49
		Average	0.55	**0.43**
	Mathematics	3	0.41	0.60
		4	0.40	0.32
		5	0.43	0.49
		6	0.72	0.87
		7	0.81	0.48
		8	0.33	0.25
		HS	0.59	0.50
		Average	0.53	**0.50**
B	ELA	3	0.48	0.43
		4	0.24	0.22
		5	0.35	0.41
		6	0.43	0.66
		7	0.25	0.57
		8	0.32	0.39
		HS	0.39	0.34
		Average	0.35	**0.43**
	Mathematics	3	0.58	0.67
		4	0.83	0.86
		5	0.51	0.68
		6	0.53	0.48
		7	0.88	0.57
		8	0.76	0.51
		HS	0.89	0.78
		Average	0.71	**0.65**

continued

Path	Content Area	Grade	Multidimensionality Effect Size	
			2022	2023
C	ELA	3	0.12	0.17
		4	DNR*	0.17
		5	DNR*	0.24
		6	0.11	0.17
		7	DNR*	0.21
		8	DNR*	0.16
		HS	DNR*	0.20
		Average	0.11	**0.19**
	Mathematics	3	0.31	0.37
		4	0.52	0.40
		5	0.28	0.37
		6	0.36	0.40
		7	DNR	0.37
		8	0.42	0.49
		HS	DNR	0.45
		Average	0.38	**0.41**

** DNR = Did not reject DIMTEST null hypothesis test*

The results for 2023 displayed in Table 8-5 show that the ELA tests tend to have lower DETECT indices than the mathematics tests. Also, Path C indices tend to be lower than Paths A and B, for both ELA and mathematics. The ELA Path C tests had the lowest indices and were the only set of tests that consistently displayed either non-rejection of the DIMTEST statistic or very low DETECT indices (weak or very weak violations of LI). For mathematics, the Path C tests also tended to be lower than for Paths A or B but displayed moderate DETECT indices. Summarizing the remaining results in Table 8-5, the DETECT indices for Paths A and B for both ELA and mathematics were predominantly (22 out of 28 cases) at a moderate level.

For comparison purposes, Table 8-5 also provides the results from last year, 2021–22. The two sets of results are mostly consistent with each other. The two strong trends for 2023 (ELA having lower indices than mathematics; Path C having lower indices than Paths A and B) were also present in 2020–21. Moreover, the detailed 2023 tendencies described were also similar to what occurred in 2021–22.

Next, an investigation was conducted to identify the possible source(s) of the LID that could help explain the DIMTEST and DETECT results. Hence, how DETECT divided the tests into clusters was investigated to see if there were any discernable patterns with respect to known substantive item characteristics. Prior to 2017–18, R9-stringers were included in the data, and a strong and consistent pattern was found related to the answer keys of the items—for tests administered on Paths A and B, the placement of the correct-response key option was a strong indicator of the cluster membership of nearly every multiple-choice item. In other words, nearly all the multiple-choice items fell into three clusters, where one cluster was dominated by items with a key of “A” (the first option), another was dominated by items with a key of “B” (the middle option, when it was available), and the third was dominated by items with a key of “C” (the last option). Note that most multiple-choice items had three answer-choice options, and the few items that did not have three options were items that had only two options, which were labeled “A” and “C” for cluster-membership labeling purposes.

The removal of the R9-stringers from the data in 2017–18 (and in all years since then) did not eliminate the key-option clustering, but it did greatly weaken its effect. This same pattern again occurred in 2023. Specifically, for 6 out of 7 grade levels for mathematics and for six out of seven for ELA, Paths A and B had at least one cluster related to one of the three keys. For Path C, there were far fewer key-related clusters. For ELA, only one of the tests had any clusters related to the keys, and for mathematics, only

two of the seven Path C DETECT analyses showed strong evidence of a key-related cluster. Overall, these results were very similar to those of past years.

For the mathematics tests, 17 of the paths also included four or more constructed-response (CR) items, and these items all clustered together for 15 of the cases. These results suggest that the CR items may be adding a weak secondary dimension to the tests. This is a finding that has often been noted in dimensionality analyses we have done on general education assessments and is generally not of concern because such items are, if anything, purer measures of the constructs of interest as compared to multiple-choice items.

For the ELA tests, the operational items also included a writing prompt, which had three trait scores. For Path A, every test showed strong evidence of the writing prompt traits clustering together within a single cluster. For Path B, the tests for grades 3 to 7 showed strong evidence for this clustering. For Path C, only grades 3 to 6 showed strong evidence for this clustering. These results indicate that the writing prompt traits tend to display evidence of dimensionality distinctiveness relative to the rest of the test, especially for Paths A and B, and especially for the lower grade levels. These results are similar to what was found in all past years, including 2018–19, the first year that writing prompts were operationally administered. This finding led to proactive measures that were taken at that time to deal with the dimensionality distinctiveness of the writing prompt traits. Specifically, the non-writing prompt items were calibrated first and then the writing prompt traits were calibrated while holding the psychometric model for the rest of test fixed. This process resulted in maintaining the same scale as was set prior to the introduction of the writing prompts while also allowing the writing prompt traits to contribute in an appropriate manner to that scale.

As in the past years, the dimensionality analysis results for Paths A and B continue to indicate a violation of local independence having to do with how some student scores are related to the placement of the correct response options. Due to the limitation that level one items offer only two answer choices (A or B), there is an increased likelihood of having consecutive items with the correct answer as A for example, despite diligent efforts in test construction to mitigate such situations. The violations of local independence are greatly reduced because of the removal of the R9-stringers from the datasets. In general, it is important that violations of local independence be understood, monitored, and controlled on tests. In particular, we understand the violation of the local independence issue is caused by R9 stringers, and actively monitor this issue by identifying stringers in all MSAA tests. By removing the R9 stringers from the calibration sample, the issue is being controlled and monitored. We also established test construction guidelines to reduce repetitive answer key distribution to prevent the test from having too many items in a row with the same answer key. The violations of local independence that are related to the ordering of the correct-response option in selected-response items are a phenomenon that will continue to require close study.

Chapter 9. Item Response Theory Scaling and Equating

This chapter describes the procedures used to calibrate, equate, and scale the 2023 MSAA. 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, and comparison of lookup tables to the previous year's lookup tables).

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), 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 s location on a latent ability continuum, in other words student 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 tests, the two-parameter logistic (2PL) model was used to estimate the ICC for dichotomous items, and the graded-response model (GRM) was used for polytomous items (Nering & Ostini, 2010). 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.

In the GRM for polytomous items, an item is scored in $k+1$ graded categories that can be viewed as a set of k dichotomies. At each point of dichotomization (i.e., at each threshold), a two-parameter model can be used to model the probability that a student's response falls at or above a particular ordered category, given θ. This implies that a polytomous item with $k+1$ categories can be characterized by k item category threshold curves (ICTCs) of the two-parameter logistic form:

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

where
U indexes the scored response on an item,
i indexes the items,
j indexes students,
k indexes threshold,
θ is the student ability,
α represents item discrimination,
b represents item difficulty,
d represents an item-category threshold, and
D is a normalizing constant equal to 1.701.

After computing k ICTCs in the GRM, $k+1$ item category characteristic curves (ICCCs), which indicate the probability of responding to a particular category given θ, are derived by subtracting adjacent ICTCs:

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

where
i indexes the items,
j indexes students,
k indexes threshold,
θ is the student ability,
P_{ik} represents the probability that the score on item i falls in category k, and
P_{ik}^{*} represents the probability that the score on item i falls at or above the threshold k ($P_{i0}^{*} = 1$ and $P_{i(m+1)}^{*} = 0$).

The GRM is also commonly expressed as:

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

Finally, the item characteristic curve (ICC) for a polytomous item is computed as a weighted sum of ICCCs, where each ICCC is weighted by a score assigned to a corresponding category. The expected score for a student with a given theta is expressed as:

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

where
w_{ik} is the weighting constant and is equal to the number of score points for score category k on item i.

Note that for a dichotomously scored item, $E(U_i|\theta_j) = P_i(\theta_j)$. 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 2023 MSAA was a pre-equated assessment program, the item parameters for the 2023 operational administration came from calibrations conducted in previous years. Items previously used operationally were calibrated in the post-equating procedures that were implemented after their corresponding operational administrations. Items previously used only as field-test items were calibrated in the corresponding field-test calibration that occurred after the calibration of the operational items. No new calibrations were run for the 2023 MSAA prior to the reporting of scores. The procedures used to conduct the calibrations discussed above 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 off-scale 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. In addition, the item parameter estimates were inspected. The discrimination parameters should not be extreme in either direction (neither greater than 3 nor less than 0.35); the difficulty parameters should also not be extreme (generally between -3 and 3, and definitely between -4 and 4); and the standard error of the difficulty parameters should generally be less than 0.2. Items that violate any of these conditions are automatically marked as “Do Not Use,”

The equating set (a subset of the operational items) was then carefully chosen to represent the test as a whole, in terms of content coverage and difficulty levels, and the equating items were evaluated to ensure only items with statistically stable psychometric models were used. The equating set serves the purpose of regularly refreshing the existing item parameters after each administration to effectively counteract any potential item drift. These updated item parameters will then be integrated into the pre-equating solution for the subsequent administration. For any equating design, it is critical that rigorous procedures are implemented to monitor the quality of the equating and to check that the assumptions underlying the equating are not violated. Cognia psychometricians have conducted research studies (Hagge & Keller, 2009; Keller et al., 2008; Keller et al., 2007; Parker et al., 2009) in this regard and have developed tools to estimate equating error across years under realistic violations of the equating assumptions. The Psychometrics Department monitors well-known violations of IRT equating assumptions and uses the research to estimate their effects on the reliability and validity of the equating. Specifically, the equating data were analyzed in detail for scale drift through traditional *b*-*b* analyses. For example, in 2023, there are around 14-27 items that were removed from being used as the equating set during the post-equate verification.

The *b*-*b* analysis compared item parameters from the previous administration to the current administration *b* parameters using linear regression analysis. A standardized perpendicular difference from the regression line was calculated for each item; any item with a difference of a magnitude of 3 or greater was flagged for drift. Furthermore, special procedures were enacted during the calibration phase to check that the quality of the equating items was maintained consistently across years. Equating items that displayed lack of stability (e.g., standard error of the *b* parameters being large, inadequate model-data fit, etc.) were flagged and removed. Using this equating set, the Stocking-Lord transformation constants were calculated to determine the relationship between the off-scale calibration and the base-year scale

established in the first year of the program. The Stocking-Lord transformation was then applied to all the off-scale operational item parameters to bring them onto the base-year scale.

Next, the field-test items were 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 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.

9.3 Item Response Theory Results

The tables in Appendix I give the IRT item parameters for all the operational items on the 2023 MSAA tests by grade and content area based on their pre-equated calibrations. The statistics for the operational items are summarized in Tables 9-1 through 9-4. The mean item parameter estimates shown in the tables below are within the generally acceptable and expected ranges. For easy reference, Table 9-1 displays the means and standard deviations averaged across all dichotomously scored operational items for each grade and content area.

Table 9-1. IRT Summary Statistics for Dichotomously Scored Items

Content Area	Grade	Number of Items	*a*		*b*	
			mean	SD	mean	SD
ELA	3	70	0.85	0.28	-0.39	0.67
	4	73	0.86	0.33	-0.31	0.64
	5	72	0.91	0.42	-0.39	0.66
	6	64	1.10	0.60	-0.16	0.64
	7	64	0.91	0.33	-0.31	0.63
	8	64	1.05	0.51	-0.40	0.52
	HS	73	1.06	0.47	-0.37	0.58
Mathematics	3	69	0.82	0.26	-0.08	0.71
	4	67	0.89	0.28	0.13	0.61
	5	70	0.84	0.27	0.14	0.78
	6	68	1.00	0.25	-0.13	0.56
	7	71	0.85	0.26	-0.18	0.71
	8	70	0.87	0.27	-0.06	0.54
	HS	68	1.06	0.39	-0.05	0.50

Because the items were developed to correspond to different levels (Levels 1, 2, and 3), the item statistics have also been summarized by item level for ELA (Table 9-2 for the dichotomous items and Table 9-3 for the writing prompt traits) and for mathematics (Table 9-4).

Table 9-2. IRT Summary Statistics by Grade and Level—ELA Dichotomous Items

Grade	Level	n_items	a_mean	a_sd	b_mean	b_sd
3	Level 1	19	1.05	0.22	-1.08	0.24
	Level 2	29	0.77	0.31	-0.29	0.52
	Level 3	22	0.79	0.21	0.07	0.64
4	Level 1	22	1.08	0.34	-1.00	0.22
	Level 2	40	0.79	0.29	-0.05	0.47
	Level 3	11	0.67	0.16	0.11	0.66
5	Level 1	21	1.36	0.41	-1.11	0.22
	Level 2	37	0.71	0.25	-0.13	0.41
	Level 3	10	0.68	0.20	0.15	0.62
6	Level 1	22	1.54	0.71	-0.75	0.23
	Level 2	30	0.94	0.41	0.04	0.58
	Level 3	12	0.70	0.19	0.41	0.47
7	Level 1	17	1.19	0.35	-0.81	0.49
	Level 2	33	0.88	0.27	-0.27	0.50
	Level 3	14	0.66	0.17	0.21	0.61
8	Level 1	23	1.53	0.54	-0.96	0.21
	Level 2	18	0.81	0.27	-0.22	0.33
	Level 3	23	0.77	0.17	0.02	0.34
HS	Level 1	11	1.27	0.38	-0.88	0.21
	Level 2	24	0.80	0.23	-0.35	0.33
	Level 3	19	0.87	0.18	0.25	0.55

Table 9-3. IRT Summary Statistics by Trait and Level—ELA Writing Prompt Items

Trait	Level	Number of Items	a		b		d0		d1		b-d0		b-d1	
			Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
C	2	7	0.80	0.08	-0.15	0.19	0.68	0.11	-0.68	0.11	-0.83	0.16	0.53	0.27
	3	7	0.80	0.08	0.30	0.22	0.92	0.11	-0.92	0.11	-0.62	0.21	1.22	0.27
I	2	7	0.81	0.10	0.23	0.25	0.69	0.19	-0.69	0.19	-0.46	0.17	0.92	0.41
	3	7	0.84	0.18	1.41	0.46	0.96	0.24	-0.96	0.24	0.45	0.59	2.36	0.43
O	2	7	0.76	0.08	0.35	0.31	1.19	0.14	-1.19	0.14	-0.84	0.24	1.54	0.41
	3	7	0.91	0.19	1.36	0.45	1.38	0.31	-1.38	0.31	-0.02	0.54	2.74	0.56

Note. C = Conventions, I = Idea Development, O = Organization.

Table 9-4. IRT Summary Statistics by Grade and Level—Mathematics

Grade	Level	Number of Items	*a*		*b*	
			Mean	SD	Mean	SD
3	1	21	0.79767	0.28097	-0.67088	0.72354
	2	37	0.86025	0.26553	0.08723	0.49316
	3	9	0.75544	0.23692	0.50402	0.58967
4	1	19	0.80108	0.16911	-0.50943	0.28123
	2	39	0.91859	0.30836	0.40297	0.48112
	3	9	0.97401	0.34072	0.27078	0.64167
5	1	24	0.89720	0.31956	-0.75669	0.31827
	2	39	0.82901	0.24661	0.60811	0.47743
	3	7	0.74723	0.19639	0.59575	0.52428
6	1	16	1.09910	0.15511	-0.84257	0.13015
	2	40	0.98790	0.28599	-0.01275	0.37798
	3	12	0.92200	0.21309	0.41179	0.51499
7	1	18	0.86289	0.22345	-1.12714	0.31641
	2	41	0.83125	0.25654	0.01536	0.41729
	3	12	0.90801	0.31311	0.56093	0.47911
8	1	15	0.93869	0.36150	-0.82824	0.32405
	2	45	0.87461	0.24327	0.08571	0.34865
	3	10	0.73924	0.18872	0.45209	0.33046
HS	1	19	1.09866	0.29801	-0.68977	0.20162
	2	41	1.11351	0.41374	0.15093	0.31667
	3	8	0.69457	0.24577	0.47572	0.29421

Results for the dichotomously scored items are as follows. As seen in Tables 9-2 and 9-4, item difficulty tends to have a positive relationship with level: as item 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 for both ELA and mathematics. To investigate these tendencies more rigorously, a one-way analysis of variance (ANOVA) was conducted on item difficulty with level as the factor. Separate ANOVAs were run for ELA and mathematics. The average Item difficulty did not differ significantly by grade level for either ELA or mathematics. Given that all item parameters were originally calibrated using a standard normal distribution (mean=0, sd=1), it's reassuring to observe the relative difficulties of the test remain consistent across grades.

The ANOVAs indicated that item level was statistically significant for both ELA and mathematics with R-squared values of 47% for ELA and 51% for mathematics. Further Tukey paired-comparison tests were also conducted across all grades. These results showed that for both ELA and mathematics, the Tukey tests indicated statistically significant differences between Level 1 and each of the other levels. For both ELA and mathematics, the Tukey comparison for Level 2 versus Level 3 was also significant averaging across all grades. In the case of grade 4 and 5 Math, the average item difficulties for Level 3 items seemed to be less challenging than those for Level 2 items.

Discussed next are the results for the polytomously scored writing prompt traits. As shown in table 9-3, for all three traits, the Level 3 traits tend to be more difficult than the Level 2 traits, but the difference is much greater for the Idea Development trait and the Organization trait than for the Conventions trait.

The IRT statistics were also summarized by different paths (Tables 9-5 and 9-6).

Table 9-5. IRT Summary Statistics by Grade by Path—ELA Dichotomous Items

Grade	Path	Number of Items	*a*		*b*	
			Mean	SD	Mean	SD
3	A	41	0.94119	0.29627	-0.70948	0.52501
	B	41	0.80480	0.26373	-0.22925	0.66857
	C	41	0.79432	0.28857	-0.40416	0.54473
4	A	41	0.89197	0.34366	-0.63411	0.54366
	B	41	0.79151	0.29599	-0.23556	0.53833
	C	41	0.70213	0.19559	-0.14128	0.63882
5	A	39	1.09324	0.44589	-0.76483	0.49684
	B	39	0.76898	0.27457	-0.32495	0.40371
	C	39	0.73975	0.22610	-0.14604	0.62192
6	A	38	1.30823	0.65895	-0.49056	0.48400
	B	38	0.86745	0.32792	0.01226	0.62016
	C	38	0.89258	0.36402	-0.03104	0.58138
7	A	38	1.02979	0.35805	-0.57704	0.48925
	B	38	0.81155	0.25611	-0.05680	0.56171
	C	38	0.82918	0.23716	-0.21480	0.42759
8	A	38	1.20861	0.59097	-0.66790	0.43932
	B	38	0.81912	0.28638	-0.21332	0.43500
	C	38	0.84809	0.28862	-0.27312	0.41359
HS	A	38	1.23750	0.57360	-0.71384	0.39348
	B	38	0.89139	0.26415	-0.42685	0.35117
	C	38	0.85360	0.24093	-0.10374	0.60556

Table 9-6. IRT Summary Statistics by Grade by Path—Mathematics

Grade	Path	Number of Items	*a*		*b*	
			Mean	SD	Mean	SD
3	A	35	0.83136	0.29228	-0.44838	0.55812
	B	35	0.77043	0.19694	0.27940	0.66839
	C	35	0.80409	0.25319	-0.12373	0.61232
4	A	35	0.85973	0.24898	-0.00906	0.48295
	B	35	0.86983	0.22242	-0.22667	0.40865
	C	35	0.89412	0.30089	0.34276	0.60211
5	A	35	0.89511	0.31304	-0.29044	0.76112
	B	35	0.75785	0.18102	0.24347	0.71300
	C	35	0.78838	0.20362	0.37030	0.69767
6	A	35	0.95291	0.25645	-0.11890	0.40037
	B	35	1.02163	0.23852	-0.41820	0.44898
	C	35	0.92986	0.22169	0.17879	0.48304
7	A	35	0.82680	0.26917	-0.56723	0.58383
	B	35	0.75190	0.16614	-0.26756	0.61887
	C	35	0.85038	0.25226	0.07875	0.72998
8	A	35	0.85999	0.28998	-0.31319	0.50812
	B	35	0.78766	0.14505	0.13669	0.44940
	C	35	0.85060	0.24998	-0.07659	0.45221
HS	A	35	1.04609	0.26343	-0.31209	0.40013
	B	35	1.01934	0.41619	0.17491	0.41917
	C	35	0.98478	0.32012	-0.07956	0.46741

The average item difficulty substantially increased from Path A to Path B for nearly all ELA tests and for most of the mathematics tests, as intended. Difficulty also usually increased from Path B to Path C for 3

out of 7 cases in mathematics and for 6 out of 7 cases in ELA, although in most cases the difference was small.

The TCCs provide a more complete picture of the various paths. TCCs display the expected (average) raw score associated with each θ_j value between -2.0 and 2.0. 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 follows:

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

Appendix J shows graphs of the TCCs and TIFs for each grade and content area.

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.

All 2023 MSAA tests used item pre-equating methodology as described in Kolen and Brennan (2014). Item pre-equating allows the raw-to-scaled score conversion to be produced before a form is administered, which in turn allows for faster reporting and turnaround times. In item pre-equating, new forms are built from a pool of pre-existing IRT-calibrated items. In addition to these operational items, new

non-operational items (field-test items) were also included on the forms. The operational items were used as a set of common items for transforming the item parameters of the non-operational items so that they would be on the same theta scale as the IRT-calibrated item pool. This allows for the item pool to be expanded continually.

However, with pre-equating, several cautions need to be taken into consideration. For example, Kolen and Brennan (2014) suggest that, to ensure that items behave the same on each administration, the items should appear in the same contexts and positions operationally as they did non-operationally. Thus, care must be taken to avoid significant shifts in position and context. Any drift must be carefully monitored and controlled to ensure comparability between forms of the test.

The item parameters for scoring the 2023 operational tests were based on post-equated calibrations conducted on past operational administrations. The raw score to scaled score lookups based on the pre-equated model for the items used in the 2023 administration are displayed in Appendix K.

Post-equating procedures are conducted after every operational administration. For any equating design, it is critical that rigorous procedures are implemented to monitor the quality of the equating and to check that the assumptions underlying the equating are not violated. The equating data are analyzed in detail for scale drift through traditional *b*-*b* analyses.

During the post-equating, item parameter estimates are placed on the base-year scale (i.e., the item bank scale) by using the method of Stocking and Lord (1983), which is based on the IRT principle of item parameter invariance. According to this principle, the equating items for both the base year and current year tests should have the same item parameters. After the item parameters for each current year's test are estimated using PARSCALE (Muraki & Bock, 2003), the Stocking and Lord method is employed to find the linear transformation (slope and intercept) that adjusts the equating items' parameter estimates such that the current year's test characteristic curve (TCC) for the equating items is as close as possible to that of the base year's tests.

In addition, the calibrated and equated parameters are evaluated to further investigate drift at both the item and test levels. At the item level, the individual item parameters are compared and investigated, and at the test level, the TCC, test information function (TIF), and raw score cuts are compared. Finally, the item parameters resulting from this process are updated in the item bank, and these updated parameters are used in field-test calibrations and in future test form development. Given that MSAA is a stage-adapted assessment and MSAA students demonstrated a wide range of performance estimates,iItems that are part of stage 2 are taken by a subset of students. This creates additional challenges to use a universal measure to indicate whether the pre-equated solution is sufficient. Those additional challenges include a restricted range of MSAA students taking items in stage 2. Therefore, a different approach for the post-equate evaluation was implemented where the field-test item statistics were used as the starting point of an item estimates and then evaluate the item fit plots every year to determine if the field-test parameters are still appropriate or the post-equated item parameters are more appropriate.

9.5 Reported Scale Scores

Because the θ scale used in IRT calibrations is not readily understood by most stakeholders, reporting scales were developed for MSAA. 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 originally fixed at the August 2015 standard setting to be 1240 for

each grade level, but some of the scale score cuts, including some of the second cuts, were adjusted during the July 2018 standards validation, as evidenced in Table 9-8.

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 2021 MSAA 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 2021 MSAA tests can be expressed in raw scores (where in this case a "raw score" is a score in the theta metric of logits) or scale scores (a 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 (nor across Paths A, B, and C) because they are affected by differences in group ability and/or 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. Scale scores are calculated using the following linear equation:

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

where
m is the slope, and
b is the intercept.

For MSAA, the base-form operational scale was set so that the theta corresponding to the proficient cut from the August 2015 standard setting was transformed to a scale score of 1240, and so that the standard deviation of the scale scores in the base-year was 15. The lowest obtainable scale score (LOSS) was set at 1200, and the highest obtainable scale score (HOSS) was set at 1290. A separate linear transformation is used for each grade and content-area combination. Because only one point within the θ scale score space and the standard deviation of the scale was fixed, 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. When the standards validation was conducted in July 2018, the transformation constants established in the base year were not modified, but some of the theta cuts were modified, including some of the Level 2/Level 3 cuts (i.e., the proficient cuts). Thus, scale score value for the proficient cut is no longer equal to 1240 for some tests (as seen in Table 9-8).

Table 9-7 shows the slope and intercept values used to calculate the scale scores for each content area and grade. Note that the values in Table 9-7 will not change unless the standards are reset.

Table 9-7. Scale Score Slope and Intercept by Content Area and Grade

Content Area	Grade	Slope	Intercept
ELA	3	11.7202	1242.0537
	4	12.0593	1240.0910
	5	12.4236	1241.6149
	6	12.3522	1237.8126
	7	12.2964	1242.4332
	8	12.6082	1239.4570
	HS	11.4922	1244.2240
Mathematics	3	13.0552	1243.6651
	4	13.1002	1239.8674
	5	13.0769	1241.4102
	6	12.8203	1241.2532
	7	12.9093	1243.2438
	8	13.0213	1242.3583
	HS	12.9897	1242.4799

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

9.6 MSAA Performance Levels, Cut Scores, and Standards Validation

Cut scores for MSAA in ELA and mathematics were originally set in a standard setting process that took place in August 2015. Details of the standard setting procedures can be found in the standard setting report (Measured Progress, 2015). In July 2018, Cognia and the MSAA Psychometric Subcommittee conducted a standards validation. Standards validation does not change the scale; its purpose is only to determine whether adjustments to the cut scores are needed.

The standards validation process for the 2018 MSAA was necessary to ensure that cut scores, set in 2015 for the assessments, continue to provide valid interpretation of ELA and mathematics performance using the Performance-Level Descriptors (PLDs). The standards for both ELA and mathematics were vertically articulated, using 2017 performance data, to update the performance standards and provide a coherent basis for interpreting 2018 scores and performance, and in preparation for validating the ELA standards further. No additional steps were necessary to validate the mathematics performance standards. The validation process for the ELA performance standards was necessitated by the addition of the open-response writing prompt scores to the existing ELA score scale in 2018.

A complete description of the standards articulation and validation processes appears in the 2018 *MSAA Standards Validation Report.* (See Appendix M of the 2018 *MSAA Technical Report* located online here: https://cms.azed.gov/home/GetDocumentFile?id=5cb0b3b61dcb2511e88cfef7).

Final cut scores, after mathematics and ELA vertical articulation and ELA standards validation for the 2017-18 MSAA, appear in Table 9-8.

Table 9-8. Cut Scores on the Theta Metric and Reporting Scale

Content Area	Grade	Theta				Scale Score			
		Cut1	Cut2	Cut3	Minimum	Cut1	Cut2	Cut3	Maximum
ELA	3	-0.70318	-0.21788	0.97664	1200	1234	1240	1254	1290
	4	-0.53007	-0.00755	1.52654	1200	1234	1240	1259	1290
	5	-0.83676	-0.12999	1.15500	1200	1232	1240	1256	1290
	6	-0.63000	-0.10626	1.02714	1200	1231	1237	1251	1290
	7	-0.59215	-0.19788	0.94792	1200	1236	1240	1255	1290
	8	-0.75241	-0.15521	0.78177	1200	1230	1238	1250	1290
	HS	-0.76610	-0.41106	0.89860	1200	1236	1240	1255	1290
Mathematics	3	-0.70202	-0.16584	0.76660	1200	1235	1242	1254	1290
	4	-0.63872	-0.10438	0.81776	1200	1232	1239	1251	1290
	5	-0.75784	-0.10784	0.84805	1200	1232	1240	1253	1290
	6	-0.68276	-0.21475	0.72127	1200	1233	1239	1251	1290
	7	-0.75478	-0.25128	0.76727	1200	1234	1240	1254	1290
	8	-0.65755	-0.21950	0.62527	1200	1234	1240	1251	1290
	HS	-0.61432	-0.22940	0.54044	1200	1235	1240	1250	1290

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

Content Area	Grade	Number of Students	Levels				Average Scale Score	SD of Scale Score
			Level 1	Level 2	Level 3	Level 4		
ELA	3	2,347	46%	15%	27%	11%	1236.48	14.73
	4	2,457	46%	17%	29%	8%	1236.44	14.82
	5	2,428	40%	26%	26%	8%	1236.01	14.02
	6	2,405	30%	25%	33%	12%	1235.61	13.21
	7	2,384	35%	18%	34%	14%	1239.50	13.89
	8	2,476	33%	29%	26%	12%	1235.13	13.02
	HS	2,341	33%	17%	39%	11%	1239.40	13.7
Mathematics	3	2,345	37%	21%	28%	14%	1238.63	15.88
	4	2,453	26%	23%	40%	11%	1236.82	13.79
	5	2,433	21%	34%	35%	10%	1238.48	13.50
	6	2,405	23%	23%	38%	16%	1239.23	14.49
	7	2,387	30%	25%	33%	13%	1238.98	14.52
	8	2,472	31%	19%	33%	17%	1239.28	14.83
	HS	2,336	25%	24%	36%	16%	1239.00	14.00

Tables 9-10 (ELA) and 9-11 (mathematics) show the percentage of students in each performance-level category by path, along with the average and standard deviation of the scale scores for each grade/content-area combination. Note that the percentage of examinees being classified as Level 3 and Level 4 (levels of Proficient or above) increased as we move from Path A to Path C. This trend was expected due to the stage adaptive nature of the MSAA.

Table 9-10. Performance-Level Distributions by Path—ELA

Grade	Path	Number of Students	Levels				Average Scale Score	SD of Scale Score
			Level 1	Level 2	Level 3	Level 4		
3	A	1,208	87%	12%	1%	0%	1226.25	10.95
	B	530	6%	40%	54%	0%	1239.97	4.02
	C	609	0%	1%	56%	43%	1253.71	8.35
4	A	1,238	87%	12%	1%	0%	1225.85	10.37
	B	746	7%	36%	56%	0%	1241.03	5.21
	C	473	0%	0%	59%	41%	1256.94	8.30
5	A	1,181	81%	19%	0%	0%	1225.77	9.74
	B	728	2%	56%	41%	0%	1239.13	5.12
	C	519	0%	0%	62%	38%	1254.92	7.61
6	A	866	81%	17%	1%	0%	1223.57	11.03
	B	751	3%	58%	39%	0%	1236.02	3.77
	C	788	0%	2%	61%	37%	1248.45	7.81
7	A	818	87%	12%	1%	0%	1226.23	11.59
	B	750	16%	43%	41%	0%	1239.29	4.30
	C	816	0%	0%	59%	40%	1253.01	7.28
8	A	1,133	71%	28%	1%	0%	1225.50	9.86
	B	637	3%	60%	37%	0%	1236.43	4.01
	C	706	0%	3%	56%	41%	1249.41	8.34
HS	A	838	82%	15%	3%	0%	1227.69	12.87
	B	818	10%	31%	59%	0%	1240.71	4.01
	C	628	0%	0%	61%	39%	1253.60	7.00

Table 9-11. Performance-Level Distributions by Path—Mathematics

Grade	Path	Number of Students	Levels				Average Scale Score	SD of Scale Score
			Level 1	Level 2	Level 3	Level 4		
3	A	959	86%	14%	0%	0%	1225.20	12.45
	B	663	6%	54%	40%	0%	1240.03	4.19
	C	723	0%	1%	53%	46%	1255.15	9.11
4	A	729	81%	18%	1%	0%	1221.86	12.40
	B	1,000	5%	45%	50%	0%	1238.19	4.28
	C	724	0%	0%	64%	36%	1250.01	7.86
5	A	802	63%	37%	0%	0%	1225.52	11.98
	B	941	1%	55%	44%	0%	1239.91	4.32
	C	690	0%	2%	62%	35%	1251.58	9.02
6	A	711	76%	24%	0%	0%	1224.10	12.75
	B	948	2%	39%	59%	0%	1239.59	3.93
	C	746	0%	1%	48%	50%	1253.19	9.46
7	A	753	84%	15%	1%	0%	1224.97	11.63
	B	853	9%	55%	37%	0%	1238.33	4.21
	C	781	0%	2%	59%	39%	1253.21	10.14
8	A	823	84%	15%	1%	0%	1225.13	11.71
	B	752	10%	44%	46%	0%	1238.80	3.77
	C	897	0%	2%	52%	46%	1252.65	10.28
HS	A	705	73%	26%	1%	0%	1225.49	13.65
	B	725	8%	47%	45%	0%	1239.02	3.46
	C	850	0%	2%	56%	42%	1250.42	8.69

Chapter 10. Reliability

Although the psychometric characteristics of individual items performance are an important focus for evaluation, a complete evaluation of an assessment must also address the way items function together. 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 being either higher or lower than his or her true 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 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 a number of ways to estimate test reliability. The most common method is Cronbach's alpha (Cronback, 1951), which is premised on a design in which all students for a given assessment were administered the same fixed set of items. For the 2023 MSAA, there were three different paths (A, B, and C), each of which essentially corresponded to a different test form. Even though Cronbach's alpha could be applied to each form separately, this would not be ideal for two reasons. First, the ability distributions for the three forms are very different from each other by design—essentially the standard deviation for any one form is much smaller than the standard deviation for the whole population, and the mean increases from Paths A to B to C. The resulting restriction of ability range causes severe underestimation of reliability for each path. Second, a single measure of reliability for each grade-level assessment is preferable to three values. Thus, an item response theory (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 multistage test in ELA and mathematics, using the scale scores (and their standard errors) for all the students across all three paths. The reliability of a test can also be evaluated by simply examining directly 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.

Tables 10-1 and 10-2 present descriptive scale score statistics, IRT-based reliability, and mean scale score CSEMs for ELA and mathematics by grade. (Statistics are based on operational items, which counted toward students' reported scores only.) As shown in the tables, all the values reached levels associated with adequate reliability (0.80 or higher).

Table 10-1. IRT Marginal Reliability by Grade—ELA

Grade	Min	Max	Mean	SD	IRT Marginal Reliability	Mean Scaled CSEM
3	1200	1289	1240.31	11.17	0.92	3.04
4	1200	1290	1240.39	12.37	0.93	3.16
5	1207	1290	1239.74	11.73	0.92	3.06
6	1203	1290	1239.20	9.70	0.90	2.91
7	1203	1290	1242.67	9.81	0.90	3.04
8	1200	1290	1238.49	10.26	0.91	2.92
HS	1207	1287	1243.40	9.51	0.91	2.65

Table 10-2. IRT Marginal Reliability by Grade—Mathematics

Grade	Min	Max	Mean	SD	IRT Marginal Reliability	Mean Scaled CSEM
3	1200	1290	1242.41	10.86	0.87	3.76
4	1203	1290	1239.02	10.02	0.88	3.37
5	1200	1290	1241.92	10.42	0.86	3.73
6	1209	1290	1241.66	11.04	0.89	3.22
7	1206	1290	1241.70	12.13	0.89	3.75
8	1201	1290	1241.78	10.85	0.85	3.70
HS	1200	1290	1242.07	8.95	0.84	3.07

10.2 Subgroup Reliability

The reliability coefficients discussed in the previous section were based on all students who took a particular 2023 MSAA 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, LEP status, socioeconomic status, migrant status, and various disability groups. Appendix M 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 M 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 when there is a restriction of range (Draper & Smith, 1998) as occurs for subgroups with little variability. Second, there is no industry standard to interpret the strength of a reliability coefficient, especially when the population of interest is a single subgroup.

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 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 methodologies 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 2023 MSAA because it can be easily applied to data that is scored in the IRT theta metric or any linear transformation of this metric, such as the MSAA 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. Thus, the Rudner method can be used to provide a single index for a multistage test, whereas an observed score method would need to be separately applied to each path of a multistage test.

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 multistage tests would fall into classification *i* (where *i* = 1 to 4) and whose observed scale score on the second hypothetical parallel multistage 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.

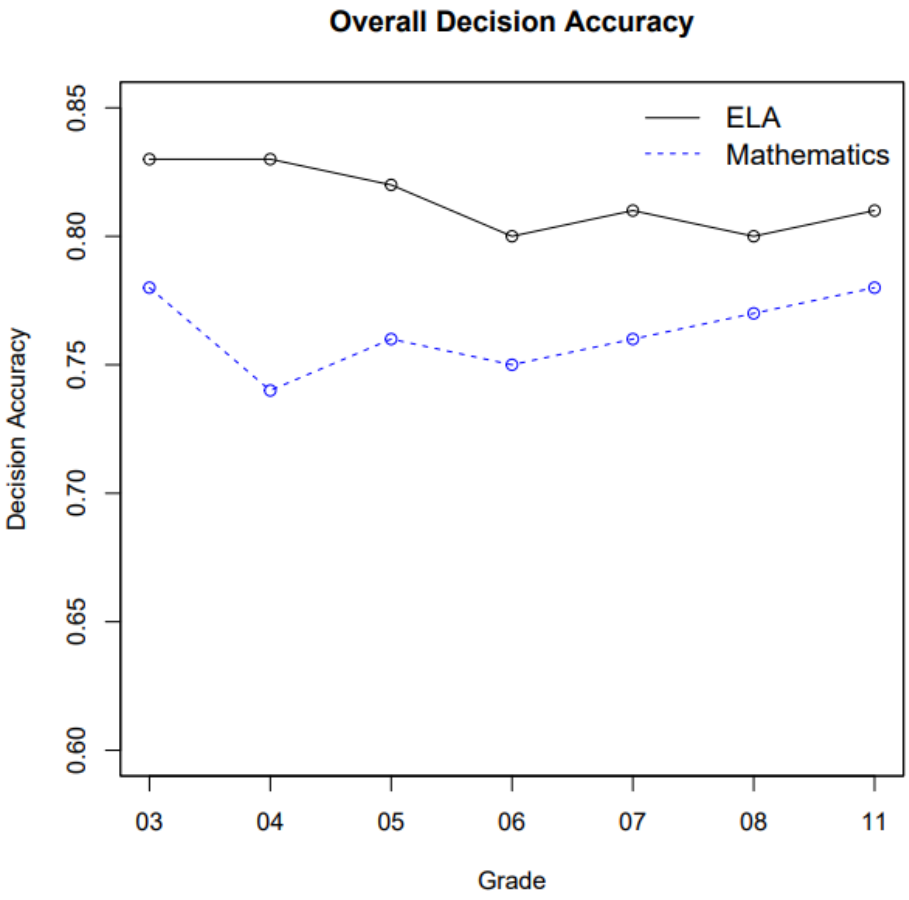
Figure 10-1 shows the overall decision accuracy for ELA and mathematics by grade level. ELA overall has higher decision accuracy (above 0.8) than math (above 0.7). Across all grades, more than 80% and 70% of ELA and Math students would be expected to be at the same performance level again when categorized according to their observed scale score, respectively. More details on decision accuracy and consistency (DAC) are provided in Appendix N. Table N-1 in Appendix N includes overall accuracy and consistency indices, along with kappa. Accuracy and consistency values conditional on performance level are also provided in Table N-1. For these calculations, the denominator is the proportion of students

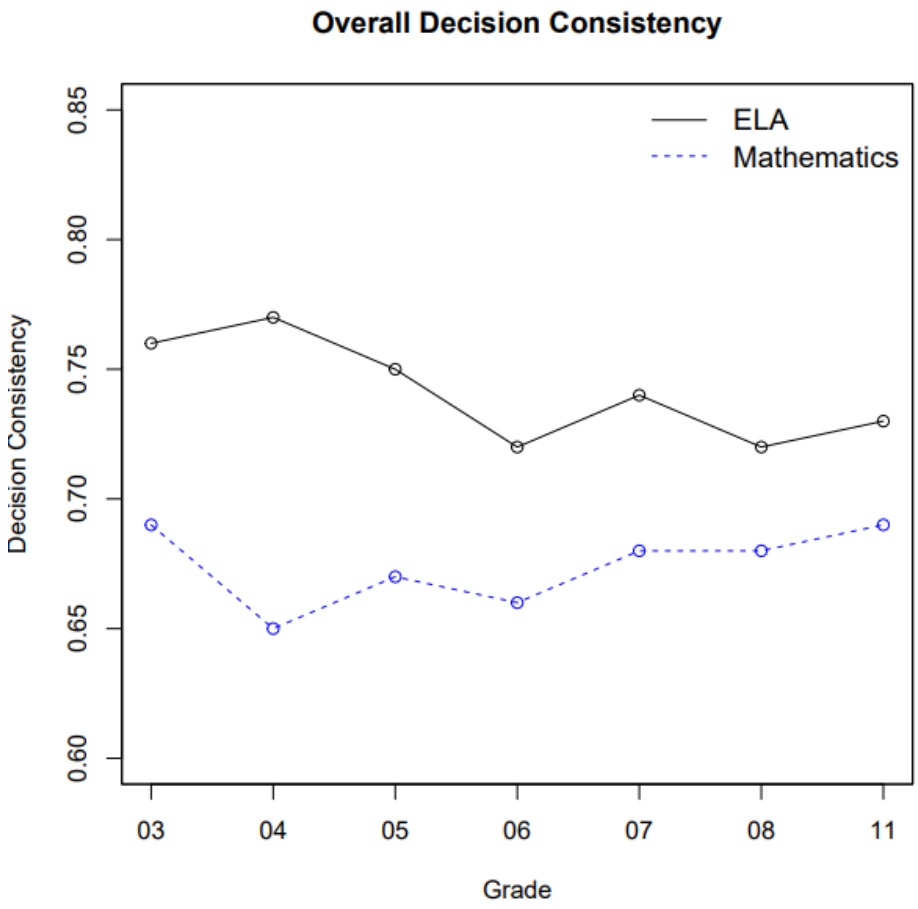
associated with a given performance level. Following is an example from Table N1, looking at Level 1 for grade 3 ELA.

- The conditional *accuracy* value was 0.91. This indicates that among the students whose *true scale scores* placed them in Level 1, 91% would be expected to be in this same level again when categorized according to their observed scale scores.
- The *consistency* value was 0.87. This indicates that among the students whose *observed scale scores* placed them in Level 1, 87% 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 accountability purposes, the primary concern is distinguishing between students who are proficient and those who are not yet proficient. For the 2023 MSAA, Table N-2 in Appendix N provides accuracy and consistency estimates at each cutpoint, as well as false positive and false negative decision rates. A false positive rate is the proportion of students whose observed scores were above the cut and whose true scores were below the cut. A false negative rate is the proportion of students whose observed scores were below the cut and whose true scores were above the cut.

Figure 10-1. Overall Decision Accuracy and Consistency by Content Area by Grade




Overall Decision Consistency
ELA
Mathematics
Decision Consistency
0.60
0.65
0.70
0.75
0.80
0.85
03
04
05
06
07
08
11
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 (SIUs—see Appendix O for a list of acronyms) of the MSAA 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 chapter uses a three-dimensional rating scale to evaluate the evidence supporting each SIU claim, aiming for a comprehensive evaluation (see Appendix P for additional details concerning the element level rating scale).

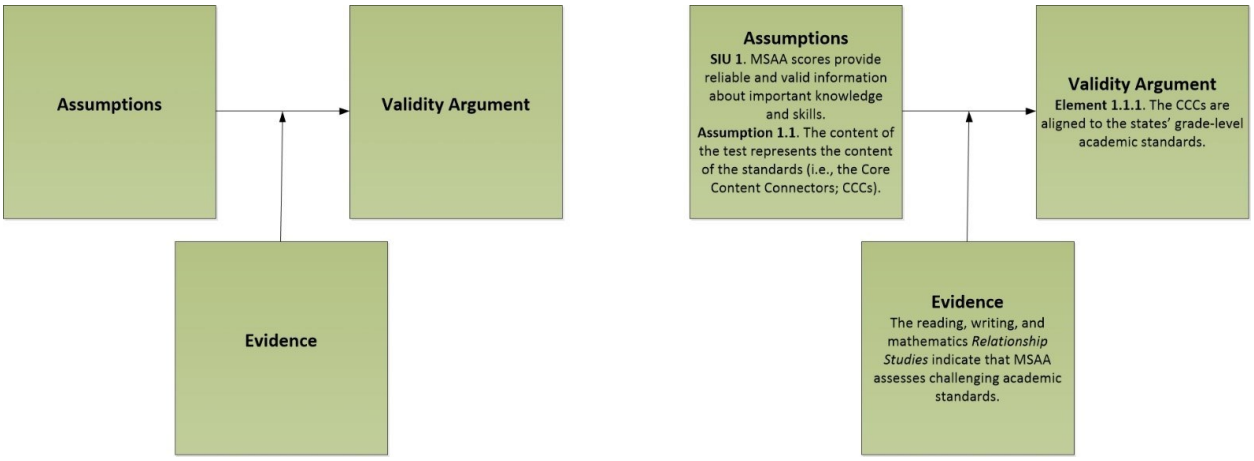
The MSAA Validity Argument Logic Model, Evidence Evaluation, and Rating Scale

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 is 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	
MSAA scores provide reliable and valid information about understanding important knowledge and skills in grade-level numeracy and literacy that students with the most significant cognitive disabilities are attaining.	
.1	The content of the test represents the content of the standards (i.e., the Core Content Connectors).
	1.1.1. MSAA content is aligned to the CCCs and grade-level standards.
	1.1.2. MSAA items are aligned to the CCCs.
	1.1.3. CCCs are aligned with the MSAA partners' content standards.
	1.1.4. MSAA items are aligned to the PLDs.
1.2	MSAA 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. Scoring rubrics focus on construct-relevant aspects of student responses.
	1.2.5. Scaffolding (information provided to vary item difficulty) does not introduce irrelevant variation to the construct.
	1.2.6. Item rendering (i.e., how items are presented in the testing platform) does not interfere with student access to test content.
	1.2.7. Test Administration Platform does not interfere with student interaction with test content.
	1.2.8. Items are free of bias and sensitive issues.

continued

Necessary Assumptions	Elements That Support Assumptions
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 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 for Paths A, B, and C 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.
	1.4.6. Stage 1 covers a broad enough range of item difficulty and item cognitive complexity to route students into appropriate stage 2 tests.
	1.4.7. Routing to Stage 2 appropriately differentiates student performance across the spectrum.
	1.4.8. Stage 2 test levels are sufficiently separable and targeted toward different ranges of achievement for the MSAA students who are routed to those levels.
1.5	Item and test scoring in 2023 were implemented accurately.
	1.5.1. Machine-scored items were scored accurately.
	1.5.2. Constructed-response item scoring training and monitoring procedures met industry standards.
1.6	MSAA 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 and its results to (a) monitor trends in school performance, and (b) design professional development for teachers on how to monitor trends.	
2.1	MSAA scores enable teachers and school, district, and state leaders to monitor trends in student proficiency.
	2.1.1. MSAA 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 scores and proficiency level categorizations of groups of students are adequately reliable to enable monitoring of grade-level performance and student cohort performance.
2.2	MSAA results are used to design professional development for teachers.
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 and its results to better integrate assessment with their instructional planning.

continued

Necessary Assumptions	Elements That Support Assumptions
	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 scores and other information for instructional planning.
	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 scores and other information useful for understanding what their child knows and can do.
	4.1.1. Parents understand and interpret correctly 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.
4.2	Parents find MSAA scores and other information useful for understanding their child's progress from year to year.
	4.2.1. Parents understand and interpret correctly MSAA scores and other information to understand their child's progress from year to year.
	4.2.2. Parents use MSAA scores and other information appropriately to understand their child's progress from year to year 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 Core Content Connectors, we lack direct evidence regarding the connection between these content standards 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 Score Interpretation

MSAA scores provide reliable and valid information about understanding important grade-level numeracy and literacy knowledge and skills attained by students with the most significant cognitive disabilities.

Primary Intended MSAA Score Uses

- Schools and districts use the MSAA and its results to (a) monitor trends in student performance and (b) design professional development for teachers.
- Teachers use the MSAA and its results to better integrate assessment with their instructional planning.
- 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.

11.1 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.

Assumption 1.1. The content of the test represents the content of the standards (i.e., the Core Content Connectors).

1.1.1. MSAA content is aligned to the CCCs and grade-level standards.
1.1.2. MSAA items are aligned to the CCCs.
1.1.3. CCCs are aligned with the MSAA partners' content standards.
1.1.4. MSAA items are aligned to the PLDs.

The evidence supporting the alignment of the Multi-State Alternate Assessment (MSAA) with academic content standards, as discussed in the 2023 MSAA ELA and Mathematics Technical Report, strongly supports the following elements: strongly supports the following elements:

- The Core Content Connectors are aligned to the partners' grade-level academic content standards. The evidence indicates that the MSAA CCCs are strongly aligned with the Common Core State Standards (CCSS) based on content centrality, performance centrality, cognitive complexity, and depth of knowledge levels.
- The 2023 MSAA items are aligned to the Core Content Connectors. The operational items of the MSAA were designed to assess the knowledge and skills of students with significant cognitive disabilities, and item specifications and development/review processes ensure alignment with the CCCs.

- The MSAA partners have confirmed alignment between the MSAA Core Content Connectors and each partner's academic content standards. Membership in MSAA requires adoption of the academic content standards assessed on the MSAA and alignment is further supported by partner review of item specifications and item content as well as content reviews from educators representing partners in item review committees.
- The 2023 operational MSAA items are aligned to the MSAA performance-level descriptors (PLDs). Panelists rated item groups based on the alignment of knowledge, skills, and abilities (KSAs) in the items with the PLDs, with acceptable overlap of KSAs found. The item specifications also provide guidelines for alignment with the PLDs, although some challenges with more recent items and PLDs were identified. Overall, the existing evidence supports this element, but further validation may be warranted through a follow-up study.

In summary, the evidence from alignment studies conducted by the NCSC strongly supports the alignment of MSAA with academic content standards, CCCs, and PLDs, with some possible challenges identified for more recent items and PLDs that may require further validation.

Relevance: Evidence is highly relevant

Completeness: Evidence is Moderate to Substantial

Overall Support: Existing evidence strongly supports the assumption.

Assumption 1.2. MSAA 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.
1.2.3. Appropriate accommodations are provided to meet student needs.
1.2.4. Scoring rubrics focus on construct-relevant aspects of student responses.
1.2.5. Scaffolding is not a source of construct-irrelevant variance.
1.2.6. Item rendering does not interfere with student access to test content.
1.2.7. Platform does not interfere with student interaction with test content.
1.2.8. Items are free of bias and sensitive issues.

The evidence supporting the alignment of the Multi-State Alternate Assessment (MSAA) with academic content standards, as discussed in the *2015 NCSC Operational Assessment Technical Manual* and *2023 MSAA ELA and Mathematics Technical Report*, strongly supports the following elements:

- The Core Content Connectors are aligned to the states' grade-level academic content standards. The evidence indicates that the MSAA CCCs are strongly aligned with the Common Core State Standards (CCSS) based on content centrality, performance centrality, cognitive complexity, and depth of knowledge levels.
- The 2023 MSAA items are aligned to the Core Content Connectors. The operational items of the MSAA were designed to assess the knowledge and skills of students with significant cognitive disabilities, and item specifications and development/review processes ensure alignment with the CCCs.

- The MSAA partners have confirmed alignment between the MSAA Core Content Connectors and each partner's academic content standards. Membership in MSAA requires adoption of the academic content standards assessed on the MSAA, and the evidence supports this alignment.
- The 2023 operational MSAA items are aligned to the MSAA performance-level descriptors (PLDs). Panelists rated item groups based on the alignment of knowledge, skills, and abilities (KSAs) in the items with the PLDs, with acceptable overlap of KSAs found. The item specifications also provide guidelines for alignment with the PLDs, although some challenges with more recent items and PLDs were identified. Overall, the existing evidence supports this element, but further validation may be warranted through a follow-up study.
- MSAA test development process aims to maximize accessibility for all students by incorporating permissible accommodations (such as presentation and communication modes) and integrating accessibility features into the system, all guided by comprehensive accessibility standards. Stringers, a relevant learner trait in this context, might lead to repetitive responses in some students. To mitigate this, new item types have been introduced in the hope of diversifying their answers.

In summary, the evidence from alignment studies conducted by the NCSC strongly supports the alignment of MSAA with academic content standards, CCCs, and PLDs, with some possible challenges identified for more recent items and PLDs that may require further validation.

Relevance: Evidence is highly relevant

Completeness: Evidence is Moderate to Substantial

Overall Support: Existing evidence strongly 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 properly.
1.3.2. Test security concerns and breaches were limited.

In summary, the MSAA assessment program ensures that Test Administrators and School and District Coordinators are trained and perform their roles appropriately, with evidence including mandatory training, online modules, and observation checklists. Results from observations indicate understanding and appropriate performance of roles. Test security concerns are addressed through irregularity reports, with no significant problems noted. The evidence supports the claim that test security concerns were limited. More specifically,

- The assessment program ensures that Test Administrators and School and District Coordinators understand and perform their roles appropriately, with evidence provided through mandatory test administration training, online modules, supporting documents, and observation checklists. Test Administrators and Coordinators received training through various resources, completed a final quiz, and results from observations indicated that they understood and performed their roles appropriately.
- Test security concerns are minimal, as evidenced by irregularity reports filed by Test Administrators and District Test Coordinators. The evidence shows no significant problems related to disruptions or suspicious activity, and no patterns of responses that may indicate further investigation were noted. The evidence provided is moderately complete and supports the claim that test security concerns were limited.

Relevance: Evidence is highly relevant

Completeness: Evidence is limited

Overall Support: Existing evidence strongly supports the assumption.

Assumption 1.4. Test scores on the 2023 MSAA 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 support intended interpretations about all students who take the MSAA.
1.4.3. Test characteristics for Paths A, B, and C 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.
1.4.6. Stage 1 covers a broad enough range of item difficulty and item cognitive complexity to route students into appropriate stage 2 tests.
1.4.7. Routing into the stage 2 test level is appropriate for students.
1.4.8. Stage 2 test levels are sufficiently separable and targeted toward different ranges of achievement for the MSAA students who are routed to those levels.

- The reliability of MSAA scores and categorizations into performance levels is supported by evidence of internal consistency, scaled score standard errors, and performance-level classification consistency and accuracy. Existing evidence strongly supports this element, with a possible challenge being the impact of local item dependence detected in dimensionality analysis.
- The item characteristics for the MSAA support intended interpretations about all students. Evidence includes DIF analyses, dimensionality assessment, calibration, and model fit evaluation. Existing evidence strongly supports this element, with possible challenges related to local item dependence.
- Test characteristics for paths A, B, and C support intended interpretations. Evidence includes dimensionality analysis and test information functions. The evidence is highly relevant and mostly complete, moderately supporting the claim with challenges related to cut 3 TIF values.
- Scaling of the MSAA supports intended interpretations. Evidence includes DIF analyses, dimensionality assessment, calibration, and model fit evaluation. The evidence is highly relevant and mostly complete, providing moderate to strong support with potential challenges related to local item dependence.
- Equating of MSAA test forms is supported by the quality of equating items and third-party analysis. Evidence is highly relevant but moderately complete, offering moderate to substantial support, with potential for additional results presentation.
- Element 1.4.6 focuses on the adequacy of Stage 1 in covering a wide range of item difficulty and cognitive complexity to guide students into appropriate Stage 2 test levels. Element 1.4.7 assesses the appropriateness of student routing into Stage 2, while Element 1.4.8 examines the separability and targeting of Stage 2 test levels for different achievement ranges in the MSAA. The evidence supporting these elements primarily comes from the MSAA Test Construction Process for 2023, which outlines how item and test information is tailored to each stage, including routing criteria. Additionally, performance-level distributions across different test paths (A, B, and

C) are analyzed in Section 9.6 to gauge the psychometric characteristics of the stages. If properly constructed and implemented, these stages should exhibit distinguishable and logical differences in performance-level distributions.

In summary, the evidence provided for this assumption confirms the reliability and validity of MSAA scores, emphasizing their accurate classification into performance levels. It sheds light on the rigorous assessment protocols in place, which are geared toward ensuring the accuracy and precision of MSAA scores. This includes procedures such as calibration, equating evaluations, and dimensionality analysis, all of which contribute to the overall validity of the scores.

Overall, the strong separation of item/test stage difficulties across three pathways - Low, Medium, and High - serves as strong validity evidence to the thoroughness of the entire test construction process. This separation underscores the careful consideration and attention to detail that goes into MSAA test construction, making it a reliable and valid tool for assessing students' performance.
These efforts collectively ensure that MSAA scores and their associated performance-level classification are accurate indicators of student achievement.

Relevance: Evidence is highly relevant

Completeness: Evidence is complete

Overall Support: Existing evidence strongly supports the assumption.

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

1.5.1. Machine-scored items were scored accurately.
1.5.2. Constructed-response item scoring training and monitoring procedures met industry standards.

Machine-scored items are verified for accuracy through a key validation process, while constructed-response item scoring adheres to industry standards with double-blind scoring and comprehensive procedures, all strongly supported by relevant and complete evidence without challenges.

- Machine-scored items are verified for accuracy through a key validation process, detailed in Chapter 6. This ensures that designated key responses for operational multiple-choice items are correct. The evidence is highly relevant, complete, and strongly supportive, with no challenges identified.
- Constructed-response item scoring meets industry standards, employing rigorous practices like double-blind scoring. All student responses are independently evaluated by two scorers, with about 5.5% undergoing a quality check by the Scoring Team Leader. Chapter 6.2 extensively outlines the procedures, including training, benchmarking, scorer qualifications, leadership, specific scoring rules, quality control, reports, and interrater reliability. The evidence is highly relevant, complete, and strongly supportive, with no challenges noted.

Relevance: Evidence is highly relevant

Completeness: Evidence is complete

Overall Support: Existing evidence strongly supports the assumption.

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

The evidence for this assumption demonstrates that MSAA scores align as expected with other measures of student proficiency, with strong convergent validity indicated by disattenuated correlations between 2023 MSAA ELA and mathematics scale scores in grades 3-8 and HS, which range from .84 to .88, as accepted by peer reviewers due to challenges in obtaining external correlation evidence.

Table 11-2. Correlation Table Between ELA and Mathematics Test Scores by Grade.

Grade	Correlation
3	0.87
4	0.84
5	0.85
6	0.88
7	0.85
8	0.86
11	0.88

Relevance: Evidence is moderately relevant

Completeness: Evidence is limited

Overall Support: Existing evidence provides limited support for the assumption.

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

2.1.1. MSAA 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 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 scores and external measures of student achievement and growth is as expected, compared to grade-level assessments and other alternate assessments.

While individual score reliability aligns with industry standards, there is limited evidence for aggregated scores at the school level. With strong support for classification accuracy and consistency where classification accuracy is above 0.70 across all grades (as shown in figures in Chapter 10). MSAA scores show a reasonable relationship with external measures of student achievement, particularly such measures of ELA and mathematics. Because of the challenges in obtaining external assessment data for students with significant cognitive disabilities, the evidence for this assumption is limited.

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

Assumption 2.2. MSAA 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 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 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 district/school leader surveys, is 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 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 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, is 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 scores and other information useful for understanding their child's progress from year to year.

4.2.1. Parents understand and interpret MSAA scores and other information correctly to understand their child's progress from year to year.
4.2.2. Parents use MSAA scores and other information appropriately to understand their child's progress from year to year 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, is 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

Many of the assumptions and elements, that either strongly or moderately support the primary intended score interpretation and three intended score uses of MSAA scores, are backed by solid evidence. 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 11.3 below.

Primary Score Intended Score Interpretation

MSAA scores provide reliable and valid information about understanding important grade-level numeracy and literacy knowledge and skills attained by students with the most significant cognitive disabilities. Of the six assumptions that support the intended score interpretation, three have highly relevant evidence. Furthermore, three assumptions have complete evidence, with one of them supported moderately to substantially and another having limited evidence. Moreover, four assumptions enjoy strong supportive evidence, whereas two assumptions are backed by only limited support.

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 that support intended score use 1: Two assumptions possess highly relevant evidence, and their evidence is moderately to substantially complete. Both assumptions also garner moderate overall support based on the available evidence.

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: Two assumptions feature evidence that is highly relevant; nevertheless, the completeness of this evidence is limited. Despite this, both assumptions receive limited support based on the available evidence.

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: Two assumptions feature evidence that is highly relevant; nevertheless, the completeness of this evidence is limited. Despite this, both assumptions receive limited overall support based on the available evidence.

Table 11.3 Status of Relevance, Completeness, and Overall Support of the Evidence for All Four SIUs

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 understanding 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 Core Content Connectors).	X		X			X		
1.2 MSAA 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, whether they correspond to the skills and processes required in the PLDs.	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
1.3 Test administrations in MSAA states in 2023 followed prescribed, standardized procedural requirements.	X				X			X
1.4. Test scores on the 2023 MSAA provide reliable information about student performance and accurate classifications into performance levels.	X		X			X		
1.5 Item and test scoring in 2023 were implemented accurately.	X		X			X		
1.6 MSAA 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 MSAA 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			X			X	
2.2 MSAA 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 and its results to better integrate assessment with their instructional planning.		X			X		X	
3.2 Teachers use MSAA scores and other information for planning instruction.		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 scores and other information useful for understanding what their child knows and can do.		X			X		X	
4.2. Parents find MSAA scores and other information useful for understanding their child's progress from year to year.		X			X		X	

Chapter 12. Ongoing Enhancements in the MSAA Program

This chapter provides a detailed record of ongoing improvements to the MSAA program for AY23. It covers enhancements realized in 2023 and those suggested by the Technical Advisory Committee (TAC).

12.1 MSAA Validity Study

Since 2022, MSAA partners have been collaborating closely with Cognia and the MSAA Technical Advisory Committee (TAC) to create and execute the MSAA Validity Survey. This survey aims to assess the utilization of MSAA test scores in the field, as well as the professional development opportunities provided to equip educators and school and district leaders with the skills to interpret data effectively. The validity survey is designed in two phases: the pilot phase, which serves to test the survey instrument and formulate a survey sampling plan, and the operational phase, focused on data collection.

During the May 2023 TAC meeting, Cognia presented the survey sampling plan, receiving both positive feedback and constructive suggestions from the TAC. An updated sampling plan is now incorporated into phase 2 of the survey with the incorporation of specific criteria. This update includes the categorization of schools into three groups based on size: (1) specialty schools catering to primary students with disabilities, (2) traditional schools with 10 or more students participating in the MSAA, and (3) traditional schools with fewer than 10 students.

During the MSAA Planning Meeting in July 2023, discussions surrounding the Validity Study Survey revealed that Arizona and Montana, the pilot states, gathered a 30% response rate with 43 total responses. It was recommended and agreed upon that a "state" identification question should be included in the operational survey. Several MSAA Partners emphasized the importance of the operational survey, and Maine, Bureau Operated Schools (BIE), Montana, and Arizona confirmed their participation. Cognia then provided an updated survey instrument to the Technical Advisory Committee (TAC) and partners for approval. Operational survey results are expected to be available in advance of the December TAC meeting.

12.1.1 Highlights from the Pilot

Some key highlights from the pilot results include:

- Demographics: Most 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: The survey inquired about participants' usage of ISR and summary reports, and the findings are affirming. All reports are actively utilized, with ISR being particularly prominent and frequently employed during IEP meetings. Within ISR, the student's

performance level and the language across all PLD levels are the most heavily utilized elements. In the case of summary reports, the mean scale score, and the number of enrolled students have emerged as the primary areas of focus.

- 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.

The pilot phase of the survey yielded limited information, primarily because it was not representative of the 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.

12.2 Continuous Improvement in Content Development

In 2019, recommendations from Dr. Diane Browder were received. These were based on new research regarding students with significant cognitive disabilities. Collaborating with the Item Development Subcommittee, temporary updates to item specifications were made, and new approaches were implemented in 2020-2021. After field testing in 2022, these changes were permanently incorporated into the 2023 assessment.

For administration year 2023, both ELA and Math content development and field-test items continued to utilize the new item approaches. As more items incorporating these approaches are field tested and become operational, we will have a more robust understanding of their performance, as shown at the whole test level in Figures 12.1 and 12.2.

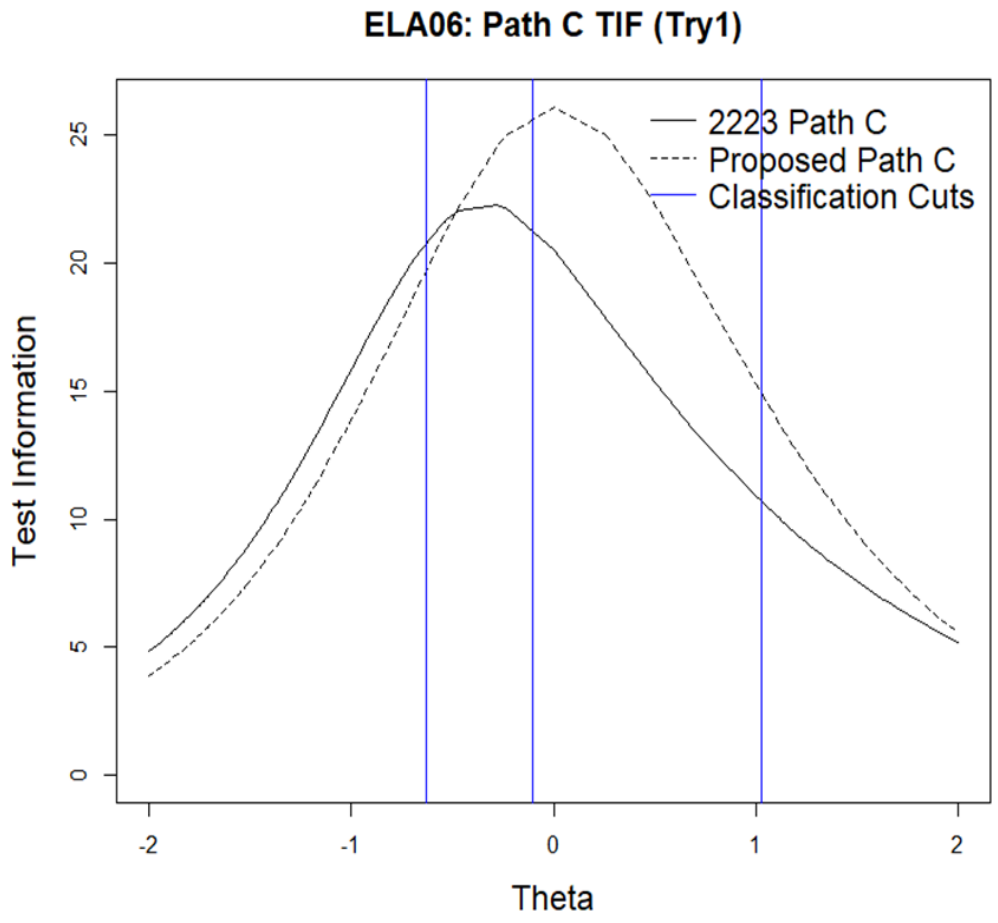
This new approach to content development has resulted in significant enhancements in the psychometric aspects of test construction, particularly in terms of test information. These improvements are especially noteworthy for the medium and high-performing MSAA students. This progress is crucial because test information directly correlates with the precision of student scores, ultimately leading to more accurate assessments.

In the case of ELA, historically, MSAA ELA items exhibited lower test information near the Level 3 cut score (cut 3) compared to Level 1 and Level 2 cut scores. To address this, the MSAA ELA development has been focused on Level 3 items and passage sets, aligning with item and passage specifications. These items feature increased content and contextual complexity while offering minimal scaffolded support, including limited graphic assistance. Consequently, this development effort has resulted in an uptick in test information at cut 3 over the past two years, furthermore, it's worth noting the substantial increase in TIF for cut 2, which is another piece of great news.

As depicted in Figure 12.1, the dotted line represents the test information for a grade 6 ELA test in the 2023-2024 school year, while the solid line represents the test information from the preceding year. The noticeable increase in test information directly translates to improved precision in student scale scores. This underscores the successful focus on MSAA test development, incorporating the Diane Browder

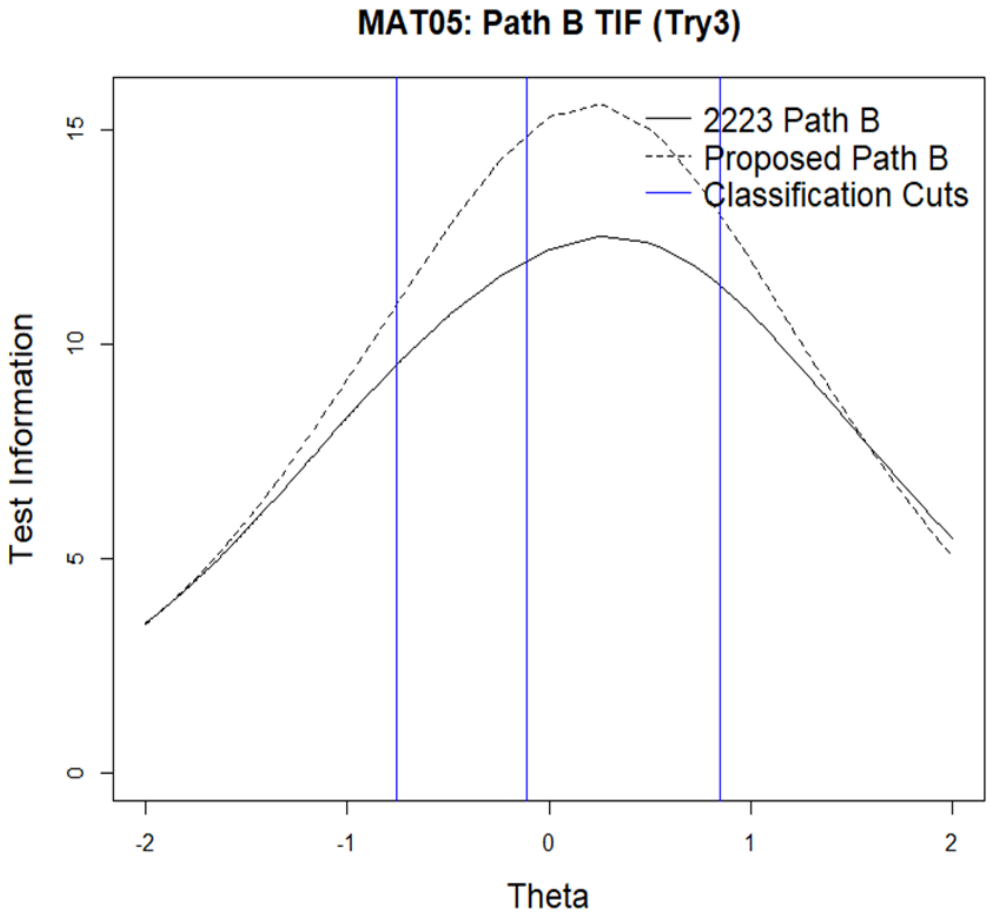
approach, deliberate collaboration between content development and psychometrics at Cognia, and improved field-test calibration procedures.

Figure 12.1 A Comparison of Grade 6 ELA Path C Test Information, 2023-2024 vs. Previous Year



In recent years, MSAA mathematics has strategically expanded its use of Constructed-Response (CR) items, while gradually incorporating newly field-tested CR items into operational assessments. This approach is supported by strong statistics from recent field-tested items, resulting in improved psychometric values. Figure 12.2 displays a Grade 5 Math test for Path B, where the dotted line represents test information. Similar to ELA, the notable spike in test information for the 2023-2024 assessment underscores the effectiveness of Math test development process.

Figure 12.2 A Comparison of Grade 5 Math Path B Test Information, 2023-2024 vs. Previous Year



12.3 MSAA Participation Rate

For students with the most significant cognitive disabilities, ESSA places a one percent threshold on their participation in a state's alternate assessment based on alternate academic achievement standards (AA-AAAS). therefore, the MSAA roster is managed by the field, typically at the state, district, or school TC level, depending on the Partner's structure. During the administration window, student rosters are regularly updated, including additions, transfers, and deletions as needed. The following information was generated from the overall test status summary report from the MSAA System, offering statistics for each Partner as well as by grade and content areas. To estimate enrollment, we calculate the total number of students who should be tested by subtracting the number of canceled assessments from the total students in the roster. The count of students who attempted the test is determined by the number of submitted tests. Table 12.1 displays MSAA participation rates by partner /entities, while Table 12.2 breaks down participation rates by grade, including content area specifics.

Table 12.1: MSAA Participation Rate by State by Subject

State or Entities	ELA			Math		
	The Total Number of Student Enrolled in the MSAA Platform	Number of students attempted the test	Participation Rate	The Total Number of Student Should be tested	Number of students attempted the test	Participation Rate
AS	32	32	100%	32	32	100%
AZ	5733	5477	96%	5735	5479	96%
BI	174	105	60%	174	102	59%
DC	473	420	89%	475	417	88%
DD	301	275	91%	301	276	92%
GU	115	106	92%	112	104	93%
ME	763	557	73%	766	563	73%
MP	63	61	97%	63	61	97%
MT	819	785	96%	821	781	95%
SD	713	632	89%	715	633	89%
TN	7885	7099	90%	7886	7086	90%
VI	66	48	73%	66	49	74%
VT	451	405	90%	452	406	90%

Table 12.2: MSAA Participation Rate by Subject and Grade

Subject	Grade	The Total Number of Students Enrolled in the MSAA Platform	Number of students attempted the test	Participation Rate
ELA	3	2410	2210	92%
	4	2526	2329	92%
	5	2519	2317	92%
	6	2540	2287	90%
	7	2537	2293	90%
	8	2605	2358	91%
	HS	2451	2208	90%
Mathematics	3	2415	2207	91%
	4	2525	2325	92%
	5	2522	2322	92%
	6	2543	2286	90%
	7	2537	2294	90%
	8	2603	2354	90%
	HS	2453	2201	90%

12.4 Longitudinal Analysis of Student Cohort Performance Trends in MSAA

As students transition from one grade to the next, it's essential to implement a progress monitoring system to assess their academic performance over time. This analysis aims to track student achievement from one grade to another, measuring changes in cohort performance. For instance, Table 12.1 illustrates the progression of student cohorts across different school years, such as Grade 3 students in cohort C27 in 2022 moving on to Grade 4 in 2023.

One way to monitor student performance over time involves aggregating performance-level data as a singular point of measurement. In this analysis, students are merged using their state student IDs, and their performance is aggregated to calculate the percentage of students scoring at Level 3 and above, providing a criterion-referenced metric for progress monitoring. In Table 12.4, it shows that 37% of Grade 3 students in cohort C27 scored at Level 3 and above in 2022, and the same cohort achieved 34% at Level 3 and above in Grade 4. With only two years' worth of data, drawing definitive conclusions may be

constrained. Nevertheless, it is imperative that we initiate continuous monitoring of MSAA student performance over time, and this serves as an initial step in our endeavors. Moreover, due to the Covid-19 pandemic, student test data for the 2019-2020 school year remains unavailable. Because of this significant disruption, the analysis intentionally omits student data from years preceding the 2019-2020 school year to ensure that the study's primary focus remains on evaluating post-Covid student performance, thus maintaining a baseline for analysis.

Table 12.3 MSAA Student Cohort Chart

Grade/year	2022	2023	2024	2025	2026	2027
3	c27	c28	c29	c30	c31	c32
4	c26	c27	c28	c29	c30	c31
5	c25	c26	c27	c28	c29	c30
6	c24	c25	c26	c27	c28	c29
7	c23	c24	c25	c26	c27	c28
8	c22	c23	c24	c25	c26	c27

Table 12.4: MSAA Percent Level 3 and Level 4 by Student Cohort

Grade	School Year					
	2021-2022	2022-2023	2023-2024	2024-2025	2025-2026	2026-2027
3	36%	38%*				
4	36%	34%				
5	37%	32%				
6	48%	43%				
7	45%	46%				
8	40%*	36%				
HS	54%*	--				

**Single year data.*
--: The gap year occurs when grade 8 students transition to grade 9, and the majority of MSAA students are not assessed until they reach grade 11.

12.5 MSAA Standards Comparison in High School

This year, with new partners joining MSAA and existing partners expressing interest in MSAA encompassing more high school grade levels (e.g., 9th grade), Cognia conducted a comparison focused on evaluating the suitability of administering the Grade 11 assessment to 9th-grade students, considering both content alignment and student performance data analysis.

12.5.1 Content alignment

The MSAA standards are rooted in the work of the National Center and State Collaborative (NCSC). The NCSC established content definitions by examining general education content, domain-specific concepts, existing research, and the Common Core State Standards (CCSS). Core Content Connectors (CCCs) were then developed to link Learning Progression Frameworks (LPFs) to the CCSS before item development.

The MSAA standards involve Core Content Connectors (CCCs) that identify grade-level core academic content, known as CCCs. Additionally, the standards include Focal Knowledge, Skills, and Abilities (FKSAs) and Essential Understandings (EUs) to define item alignment. FKSAs allow for graduated

complexity, accommodating different cognitive abilities, while EUs establish entry-level skills based on grade-specific CCCs.

In the comparison of MSAA standards in ELA and Math, Math CCCs for high school are organized by content area, while ELA CCCs are divided into two grade-level bands: 9-10 and 11-12. Although there are 11 CCCs on the MSAA ELA test blueprint for high school, slight differences in wording exist in three CCCs, reflecting higher complexity skills. Despite these differences, the FKSA and EU approach the CCC at a basic level, providing alignment flexibility. Moreover, the stage-adaptive administration of MSAA allows additional flexibility in item alignment, enhancing adaptability within the assessment process.

12.5.2 AY23 Student Performance Comparison

In AY23, some 9th-grade MSAA students took the 11th-grade ELA/Math test. The table 12.5 below categorizes students into two groups for comparison: all MSAA students (including those in both 9th and 11th grade) and specifically grade 9 students from Vermont. Grade 8 student performance is included for reference, as all eighth graders across partner states are taking the 8th-grade test.

Overall, Vermont students performed similarly to the entire MSAA student population in Grade 8 ELA/Math and high school math but demonstrated lower performance in high school ELA. The noteworthy percentage of grade 9 students scoring in Levels 3 and 4 suggests exposure to content standards and content mastery. Given that AY23 marks Vermont's first assessment year with MSAA, where both students and test administrators might not be as familiar with the test, Cognia recommends ongoing monitoring of grade 9 students in MSAA.

Table 12.5 MSAA Student Performance Comparison between Grade 9 &11

Grade	Subject	% Students in Level 1	% Students in Level 2	% Students in Level 3	% Students in Level 4	% Students in Level 3 + 4	Note:
Grade 8 (VT)	ELA	41%	24%	29%	12%	41%	As reference
Grade 8 (MSAA)	ELA	34%	29%	26%	12%	37%	As reference
Grade 9 (VT)	ELA	40%	26%	24%	10%	34%	
Grade 11 (MSAA)	ELA	35%	17%	38%	11%	49%	
Grade 8 (VT)	MAT	31%	17%	39%	13%	52%	As reference
Grade 8 (MSAA)	MAT	32%	19%	33%	17%	50%	As reference
Grade 9 (VT)	MAT	33%	20%	38%	10%	48%	
Grade 11 (MSAA)	MAT	26%	24%	35%	15%	50%	
Grade 11 (VT)	SCI	31%	18%	33%	18%	51%	As reference
Grade 11 (MSAA)	SCI	39%	24%	24%	13%	37%	As reference

References

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

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

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.

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.

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, D.C.: National Council on Measurement in Education.

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

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.

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.[3]

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

[3] *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 *Vocabulary.com* Retrieved from https://www.lexico.com/en/definition/plausibility

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.

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

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

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.

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

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

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

Wakeman, S., Lee, A., Browder, D. (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.

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—ELA

Accommodations	Grades						
	3	4	5	6	7	8	HS
LCI_Vision[1]	139	122	126	124	131	131	126
SAR_Assistive_Response_After[2]	278	216	219	238	215	215	178
SAR_No_Accomm_Needed_After[3]	466	509	543	583	600	654	707
SAR_Paper_Version_After[4]	3	0	2	1	1	0	0
SAR_Scribe_After[5]	168	142	145	129	137	122	60
SAR_Sign_Interpretation_After[6]	545	549	551	495	493	513	288

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 Frequencies—Mathematics

Accommodations	Grades						
	3	4	5	6	7	8	HS
LCI_Vision[1]	136	122	127	125	133	131	125
SAR_Assistive_Response_After[2]	279	216	220	241	216	212	175
SAR_No_Accomm_Needed_After[3]	467	509	544	584	598	656	709
SAR_Paper_Version_After[4]	3	0	2	1	1	0	0
SAR_Scribe_After[5]	169	143	145	129	137	119	61
SAR_Sign_Interpretation_After[6]	547	549	550	496	493	508	287

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-3. Accommodation Frequencies—Science

Accommodations	Grades						
	3	4	5	6	7	8	HS
LCI_Vision[1]	0	0	63	0	0	68	69
SAR_Assistive_Response_After[2]	0	0	140	0	0	141	120
SAR_No_Accomm_Needed_After[3]	0	0	214	0	0	208	207
SAR_Paper_Version_After[4]	0	0	1	0	0	0	1
SAR_Scribe_After[5]	0	0	61	0	0	61	28
SAR_Sign_Interpretation_After[6]	0	0	196	0	0	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-4. Accommodation Summary

Content Area	Grade	Number of Students Tested	
		With Accommodations	Without Accommodations
ELA	3	1,256	1,091
	4	1,298	1,159
	5	1,307	1,121
	6	1,303	1,102
	7	1,303	1,081
	8	1,390	1,086
	HS	1,219	1,065
Mathematics	3	1,256	1,089
	4	1,295	1,158
	5	1,309	1,124
	6	1,308	1,097
	7	1,302	1,085
	8	1,387	1,085
	HS	1,218	1,062
Science	5	545	497
	8	555	444
	11	458	453

APPENDIX B
TEST PARTICIPATION

Table B.1 Summary of Participation by Demographic Category—ELA

Description	Tested # Complete	Tested # No Observable Mode of Communication[1]	Total Tested	Total Percent
All Students	15,961	877	16,838	100
Female	5,389	295	5,684	34
Male	10,461	574	11,035	66
Gender Undefined	111	8	119	1
Hispanic or Latino	3,645	170	3,815	23
American Indian or Alaska Native	617	39	656	4
Asian	422	24	446	3
Black or African American	2,950	162	3,112	18
Native Hawaiian or Pacific Islander	142	3	145	1
White (non-Hispanic)	6,980	412	7,392	44
Two or More Races (non-Hispanic)	686	42	728	4
No Primary race/Ethnicity Undefined	519	25	544	3
Currently receiving LEP services	1,084	45	1,129	7
Not receiving LEP services	9,366	571	9,937	59
LEP: All Other Students	5,511	261	5,772	34
Economically Disadvantaged Students	4,444	258	4,702	28
Non-economically Disadvantaged Students	5,398	341	5,739	34
SES: All Other Students	6,119	278	6,397	38
Migrant	423	21	444	3
Non- migrant	9,312	578	9,890	59
Undefined Migrant Status	6,226	278	6,504	39
Augmentative Communication	2,952	277	3,229	19
No Augmentative Communication	12,895	588	13,483	80
Undefined Augmentative Communications	114	12	126	1
Hearing Loss	332	107	439	3
Within Normal Limits	15,629	769	16,398	97
Undefined Hearing Loss	0	1	1	0
Visual Impairment	616	284	900	5
Within Normal Limits	15,272	590	15,862	94
Undefined Visual Impairment	73	3	76	0
Sensory Stimuli Response	1,170	633	1,803	11
Follow Directions	14,791	243	15,034	89
Undefined Receptive Language	0	1	1	0
Special School	986	150	1,136	7
Regular School Self-contained	10,371	643	11,014	65
Regular School Resource Room	2,977	56	3,033	18
Regular School Primarily Self-contained	1,062	15	1,077	6
Regular School General Education	564	12	576	3
Undefined Classroom Setting	1	1	2	0
Student Communicates Primarily Through Cries	1,024	617	1,641	10
Uses Intentional Communication	3,757	193	3,950	23
Uses Symbolic Language	11,180	65	11,245	67
Undefined Expressive Communication	0	2	2	0

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

Table B-2. Summary of Participation by Demographic Category—Mathematics

Description	Tested # Complete	Tested # No Observable Mode of Communication[1]	Total Tested	Total Percent
All Students	15,954	877	16,831	100
Female	5,384	295	5,679	34
Male	10,459	574	11,033	66
Gender Undefined	111	8	119	1
Hispanic or Latino	3,644	170	3,814	23
American Indian or Alaska Native	617	39	656	4
Asian	421	24	445	3
Black or African American	2,944	162	3,106	18
Native Hawaiian or Pacific Islander	139	3	142	1
White (non-Hispanic)	6,990	412	7,402	44
Two or More Races (non-Hispanic)	679	42	721	4
No Primary race/Ethnicity Undefined	520	25	545	3
Currently receiving LEP services	1,088	45	1,133	7
Not receiving LEP services	9,351	571	9,922	59
LEP: All Other Students	5,515	261	5,776	34
Economically Disadvantaged Students	4,431	258	4,689	28
Non-economically Disadvantaged Students	5,400	341	5,741	34
SES: All Other Students	6,123	278	6,401	38
Migrant	420	21	441	3
Non- migrant	9,306	578	9,884	59
Undefined Migrant Status	6,228	278	6,506	39
Augmentative Communication	2,959	277	3,236	19
No Augmentative Communication	12,881	588	13,469	80
Undefined Augmentative Communications	114	12	126	1
Hearing Loss	333	107	440	3
Within Normal Limits	15,621	769	16,390	97
Undefined Hearing Loss	0	1	1	0
Visual Impairment	616	284	900	5
Within Normal Limits	15,265	590	15,855	94
Undefined Visual Impairment	73	3	76	0
Sensory Stimuli Response	1,161	633	1,794	11
Follow Directions	14,793	243	15,036	89
Undefined Receptive Language	0	1	1	0
Special School	982	150	1,132	7
Regular School Self-contained	10,372	643	11,015	65
Regular School Resource Room	2,976	56	3,032	18
Regular School Primarily Self-contained	1,063	15	1,078	6
Regular School General Education	560	12	572	3
Undefined Classroom Setting	1	1	2	0
Student Communicates Primarily Through Cries	1,021	617	1,638	10
Uses Intentional Communication	3,761	193	3,954	23
Uses Symbolic Language	11,172	65	11,237	67
Undefined Expressive Communication	0	2	2	0

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

Table B-3. Summary of Participation by Demographic Category—Science

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 B-4. Test Participation by Subgroup

Description	Total Tested	Invalidated	Did Not Test
ELA	16,838	205	909
Mathematics	16,831	198	924
Science	2,952	34	261

APPENDIX C
TEST DESIGN BLUEPRINTS

20-21 MSAA ELA Operational Blueprint

Notes:

- Measured Progress psychometricians have analyzed passage sets as a whole to show how well they differentiate between stages 2A, 2B, and 2C using IRT stats. For additional details about this process, please reference the Test Construction Process documentation here: T:\Contracts\MSAA\6027 - 2018\Program Management\Test Construction\TC process document
    - Goal is to move toward:
        - 2A: difficulty range-low
        - 2B: difficulty range-medium
        - 2C: difficulty range-high
- Linking passage sets may occur in Session 2A, B & C, but they will vary based on how well they differentiate based on IRT stats.
- Writing standalones are included in Session One.
- Writing Prompt-SRs (Level 1) are administered in all Session 2 versions.
- Writing Prompt-OR WP Level 2 is administered in Session 2A. Writing Prompt -OR WP Level 3 is administered in Sessions 2B and 2C.
- Reading Foundational items are added to Session 1, Form 1 for grades 3 and 4 in Field Test. Grades 3-8 & High School will have FT writing items, L1 Writing Prompts or a shortened passage set.

Item Types:

SR- Selected Response: an independent item that is not connected to any other items.

- Two-Part SR: a two-part Selected Response item in which answering one item is not dependent on answering the previous item. Students can reference the previous item without impacting their score.
- MSR- Multi-Select Response: for MSAA, this item type is a two-part Selected Response with a dependency between items, similar to an Evidence-Based Selected Response (EBSR). Students are not allowed to reference the previous item because the answer to the first item in the pair is included in the directions/stem of the second item.
- The CR writing prompts are scored based on 9 possible points, but score point 1 and 2 are collapsed for reporting purposes for a total of 6 possible points.

Blueprint Guidelines ELA

When the item pool allows, these are the blueprint guidelines that will inform test construction.

ELA Content Category	Gr 3	Gr 4	Gr 5	Gr 6	Gr 7	Gr 8	HS
Reading Literary	24-32%	24-32%	25-33%	21-30%	17-26%	17-26%	17-26%
Reading Informational	18-26%	18-26%	25-33%	26-34%	32-36%	32-36%	32-36%
Reading Vocabulary and Foundational (G3 and G4)	12-16%	12-16%	6-10%	9-11%	6-9%	6-9%	6-9%
Writing	36-38%	32-38%	31-40%	36-40%	36-40%	36-40%	36-40%

Grade 3 Targets by Standard
MSAA ELA Operational Test Blueprint Grade 3

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range	2021 Point Range
Reading: Literary	24-32%	3.RL.h1** Answer questions related to the relationship between characters, setting, events, or conflicts (e.g., characters and events, characters and conflicts, setting and conflicts) NOT 2-PART	SR, MSR one or two-part item	12-16	12-16
		3.RL.i2 Answer literal questions and refer to text to support your answer	SR		
		3.RL.k2** Determine the central message, lesson, moral, and key details of a text read aloud or information presented in diverse media and formats, including visually, quantitatively, and orally	MSR, MSR two-part		
Reading: Informational	18-26%	3.RI.h1** Identify the purpose of a variety of text features NOT 2-PART	SR	9-13	9-13
		3.RI.h4 Use illustrations (e.g., maps, photographs, diagrams, timelines) in informational texts to answer questions	SR		
		3.RI.i2 Determine the main idea of text read or read aloud or information presented in diverse media and formats, including visually, quantitatively, and orally	SR		
		3.RI.k5** Determine the main idea of a text; recount the key details and explain how they support the main idea	SR,MSR two-part		
Reading:Vocabulary and Foundational	12-16%	3.RWL.i2 Use sentence context as a clue to the meaning of a new word, phrase, or multiple meaning word	SR	6-8	6-8
		3.RWL.i1 Use context to confirm or self-correct word recognition.	SR		
Writing	36-38%	3.WI.l4 Sort evidence (e.g., graphic organizer) collected from print and/or digital sources into provided categories	SR	3-4	3-4
		3.WI.p1 Include text features (e.g., numbers, labels, diagrams, charts, graphics) to enhance clarity and meaning	SR		
		3.WL.o1 With guidance and support from adults, produce a clear, coherent, permanent product that is appropriate to the specific task, purpose (e.g., to entertain), or audience	MSR, CR	7	15
Total	**100%**			**42 Total Items**	**50 Total Points**

** Percentages are approximate with the total equaling 100%*
*** CCCs require a multipart item to assess.*
****MSR- Multi-Select Response: for MSAA, this item type is a two-part Selected Response with a dependency between items, similar to an Evidence-Based Selected Response (EBSR). Students are not allowed to reference the previous item because the answer to the first item in the pair is included in the directions/stem of the second item.*

Actual percentages by point value, not item count.

ELA Content Category	Gr 3
Reading Literary	24-32%
Reading Informational	18-26%
Reading Vocabulary and Foundational (G3 and G4)	12-16%
Writing	36-38%

Grade 4 Targets by Standard
MSAA ELA Operational Test Blueprint Grade 4

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range	2021 Point Range
Reading: Literary	24-32%	4.RL.i1 Refer to details and examples in a text when explaining what the text says explicitly	SR	12-16	12-16
		4.RL.k2** Determine the theme of a story, drama, or poem; refer to text to support answer	SR, MSR one or two-part item		
		4.RL.l1** Describe character traits (e.g., actions, deeds, dialogue, description, motivation, interactions); use details from text to support description	SR, MSR two-part		
Reading: Informational	18-26%	4.RI.h4 Use information presented visually, orally, or quantitatively (e.g., in charts, graphs, diagrams, time lines, animations, or interactive elements on Web pages) to answer questions	SR	9-13	9-13
		4.RI.i3 Determine the main idea of an informational text	SR		
		4.RI.l1** Interpret information presented visually, orally, or quantitatively (e.g., in charts, graphs, diagrams, time lines, animations, or interactive elements on Web pages) and explain how the information contributes to an understanding of the text in which it appears	SR, two-part MSR		
Reading: Vocabulary Reading Foundational	12-16%	4.RWL.i2 Use context as a clue to determine the meaning of unknown words, multiple meaning words, or words showing shades of meaning	SR	6-8	6-8
		4.RWL.j1 Use general academic and domain specific words and phrases accurately	SR		
		4.RWL.i1 Use context to confirm or self-correct word recognition.	SR		
Writing	32-38%	4.WI.q1 Provide a concluding statement or section to support the information presented	SR	3-4	3-4
		4.WI.p1 Include formatting (e.g., headings, bulleted information), illustrations, and multimedia when useful to promote understanding	SR		
		4.WL.o1 Produce a clear, coherent, permanent product that is appropriate to the specific task, purpose (e.g. to entertain), or audience	MSR, CR	5-7	13-15
Total	100%			42 Total Items	50 Total Points

** Percentages are approximate with the total equaling 100%*
*** CCCs require a multipart item to assess.*
****MSR- Multi-Select Response: for MSAA, this item type is a two-part Selected Response with a dependency between items, similar to an Evidence-Based Selected Response (EBSR). Students are not allowed to reference the previous item because the answer to the first item in the pair is included in the directions/stem of the second item.*

ELA Content Category	Gr 4
Reading Literary	24-32%
Reading Informational	18-26%
Reading Vocabulary and Foundational (G3 and G4)	12-16%
Writing	32-38%

Grade 5 Targets by Standard
MSAA ELA Operational Test Blueprint Grade 5

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range	2021 Point Range
Reading: Literary	25-33%	5.RL.b1 Refer to details and examples in a text when explaining what the text says explicitly	SR	12-16	12-16
		5.RL.c2** Summarize a text from beginning to end in a few sentences	SR, MSR single or multi-part		
		5.RL.d1 Compare characters, settings, events within a story; provide or identify specific details in the text to support the comparison	SR		
Reading: Informational	25-33%	5.RI.d5** Compare and contrast the overall structure (e.g., chronology, comparison, cause/effect, problem/solution) of events, ideas, concepts, or information in two or more texts 2 Part	SR	12-16	12-16
		5.RI.c4** Determine the main idea, and identify key details to support the main idea 2 PART	SR, MSR two-part		
		5.RI.e2 Explain how an author uses reasons and evidence to support particular points in a text	SR		
Reading: Vocabulary	6-10%	5.RWL.a2 Use context to determine the meaning of unknown or multiple meaning words or phrases	SR	3-5	3-5
Writing	31-40%	5.WI.b3 Organize ideas, concepts, and information (using definition, classification, comparison/contrast, and cause/effect)	SR	2-4	2-4
		5.WI.d1 Support a topic with relevant facts, definitions, concrete details, quotations, or other information and examples	SR		
		5.WL.h1 Produce a clear, coherent, permanent product that is appropriate to the specific task, purpose (e.g. to entertain), or audience	MSR, CR	5-7	13-15
Total	**100%**			**40 Total Items**	**48 Total Points**

** Percentages are approximate with the total equaling 100%*
*** CCCs require a multipart item to assess.*
****MSR- Multi-Select Response: for MSAA, this item type is a two-part Selected Response with a dependency between items, similar to an Evidence-Based Selected Response (EBSR). Students are not allowed to reference the previous item because the answer to the first item in the pair is included in the directions/stem of the second item.*

ELA Content Category	Gr 5
Reading Literary	25-33%
Reading Informational	25-33%
Reading Vocabulary	6-10%
Writing	31-40%

Grade 6 Targets by Standard
MSAA ELA Operational Test Blueprint Grade 6

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range	2021 Point Range
Reading: Literary	21-30%	6.RL.b2 Refer to details and examples in a text when explaining what the text says explicitly	SR	10-14	10-14
		6.RL.b3 Use specific details from the text (words, interactions, thoughts, motivations) to support inferences or conclusions about characters including how they change during the course of the story	SR		
		6.RL.c3** Summarize a text from beginning to end in a few sentences without including personal opinions 3-PART	SR, SR two-part, MSR		
Reading: Informational	26-34%	6.RI.b4 Summarize information gained from a variety of sources including media or texts	SR	12-16	12-16
		6.RI.c2** Provide a summary of the text distinct from personal opinions or judgments 2 PART	SR, MSR single or multi-part		
		6.RI.g4 Determine how key individuals, events, or ideas are elaborated or expanded on in a text	SR		
		6.RI.g6 Evaluate the claim or argument; determine if it is supported by evidence	SR		
Reading: Vocabulary	9-11%	6.RWL.a1 Use context to determine the meaning of unknown or multiple meaning words or phrases	SR	4-5	4-5
		6.RWL.c1 Use general academic and domain specific words and phrases accurately	SR		
Writing	36-40%	6.WL.c1 Organize ideas and event so that they unfold naturally	SR	3-4	3-4
		6.WL.c3 Use a variety of transition words, phrases, and clauses to convey sequence and signal shifts from one time frame or setting to another	SR		
		6.WI.h2 Produce a clear, coherent, permanent product that is appropriate to the specific task (e.g., topic), purpose (e.g., to inform), and audience (e.g., reader)	MSR, CR	6-7	14-15
Total	**100%**			**39 Total Items**	**47 Total Points**

** Percentages are approximate with the total equaling 100%*
*** CCCs require a multipart item to assess.*
****MSR- Multi-Select Response: for MSAA, this item type is a two-part Selected Response with a dependency between items, similar to an Evidence-Based Selected Response (EBSR). Students are not allowed to reference the previous item because the answer to the first item in the pair is included in the directions/stem of the second item.*

ELA Content Category	GR 6
Reading Literary	21-30%
Reading Informational	26-34%
Reading Vocabulary	9-11%
Writing	36-40%

Grade 7 Targets by Standard
MSAA ELA Operational Test Blueprint Grade 7

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range	2021 Point Range
Reading: Literary	17-26%	7.RL.i2** Use two or more pieces of textual evidence to support inferences, conclusions, or summaries of text	SR, SR two-part	8-12	8-12
		7.RL.j1 Analyze the development of the theme or central idea over the course of the text	SR		
Reading: Informational	32-36%	7.RI.j1** Use two or more pieces of evidence to support inferences, conclusions, or summaries of text	SR, SR two-part	15-17	15-17
		7.RI.j5 Analyze the interactions between individuals, events, and ideas in a text (e.g., how ideas influence individuals or events, or how individuals influence ideas or events)	SR		
		7.RI.I1** Compare/contrast how two or more authors write about the same topic	SR, SR two-part		
		7.RI.k4 Evaluate the claim or argument to determine if they are supported by evidence	SR		
Reading: Vocabulary	6-9%	7.RWL.g1 Use context as a clue to determine the meaning of a grade appropriate word or phrase	SR	3-4	3-4
Writing	36-40%	7.WL.o1 Select or provide a concluding statement or paragraph that follows from the narrated experiences or events.	SR	2-4	2-4
		7.WL.I1 Use precise words and phrases, relevant descriptive details, and sensory language to capture the action and convey experiences and events	SR		
		7.WI.o1 Produce a clear, coherent, permanent product (e.g. select/generate responses to form paragraph/essay) that is appropriate to the specific task (e.g., topic), purpose (e.g., to inform), and audience(reader)	MSR, CR	7	15
Total	**100%**			**39 Total Items**	**47 Total Points**

** Percentages are approximate with the total equaling 100%*
*** CCCs require a multipart item to assess.*
****MSR- Multi-Select Response: for MSAA, this item type is a two-part Selected Response with a dependency between items, similar to an Evidence-Based Selected Response (EBSR). Students are not allowed to reference the previous item because the answer to the first item in the pair is included in the directions/stem of the second item.*

ELA Content Category	GR 7
Reading Literary	17-26%
Reading Informational	32-36%
Reading Vocabulary	6-9%
Writing	36-40%

Grade 8 Targets by Standard
MSAA ELA Operational Test Blueprint Grade 8

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range	2021 Point Range
Reading: Literary	17-26%	8.RL.i2** Use two or more pieces of evidence to support inferences, conclusions, or summaries of text	SR, SR two-part	8-12	8-12
		8.RL.j2 Analyze the development of the theme or central idea over the course of the text including its relationship to the characters, setting, and plot	SR		
Reading: Informational	32-36%	8.RI.j1** Use two or more pieces of evidence to support inferences, conclusions, or summaries of text 2 PART	SR, SR two-part	15-17	15-17
		8.RI.l1 Analyze a case in which two or more texts provide conflicting information on the same topic and identify where the texts disagree on matters of fact or interpretation	SR		
		8.RI.k2 Determine how the information in each section contribute to the whole or to the development of ideas	SR		
		8.RI.k4 Identify an argument or claim that the author makes	SR		
Reading: Vocabulary	6-9%	8.RWL.g1 Use context as a clue to the meaning of a grade-appropriate word or phrase	SR	3-4	3-4
		8.RWL.i1 Use general academic and domain specific words and phrases accurately	SR		
Writing	36-40%	8.WP.k2 Create an organizational structure in which ideas are logically grouped to support the writer's claim	SR	2-4	2-4
		8.WP.j1 Gather relevant information (e.g., highlight in text, quote or paraphrase from text or discussion) from print and/or digital sources	SR		
		8.WI.o1 Produce a clear, coherent, permanent product (e.g. select/generate responses to form paragraph/essay) that is appropriate to the specific task (e.g., topic), purpose (e.g., to inform), and audience (e.g., reader)	MSR, CR	7	15
Total	**100%**			**39 Total Items**	**47 Total Points**

** Percentages are approximate with the total equaling 100%*
*** CCCs require a multipart item to assess.*
****MSR- Multi-Select Response: for MSAA, this item type is a two-part Selected Response with a dependency between items, similar to an Evidence-Based Selected Response (EBSR). Students are not allowed to reference the previous item because the answer to the first item in the pair is included in the directions/stem of the second item.*

ELA Content Category	Gr 8
Reading Literary	17-26%
Reading Informational	32-36%
Reading Vocabulary	6-9%
Writing	36-40%

High School Targets by Standard
MSAA ELA Operational Test Blueprint--HS

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range	2021 Point Range
Reading: Literary	17-26%	1112.RL.b1** Use two or more pieces of evidence to support inferences, conclusions, or summaries of the plot, purpose, or theme within a text	SR, SR two-part	8-12	8-12
		1112.RL.d1 Analyze how an author's choices concerning how to structure specific parts of a text (e.g., the choice of where to begin or end a story, the choice to provide a comedic or tragic resolution) contribute to its overall structure and meaning	SR		
Reading: Informational	32-36%	1112.RI.b1** Use two or more pieces of evidence to support inferences, conclusions, or summaries or text	SR, SR two-part	15-17	15-17
		1112.RI.b5** Determine how key details support the development of the central idea of a text	SR, SR two-part, MSR		
		1112.RI.d1 Determine the author's point of view or purpose in a text	SR		
		1112.RI.e1 Integrate and evaluate multiple sources of information presented in different media or formats (e.g., visually, quantitatively) as well as in words in order to address a question or solve a problem	SR		
Reading: Vocabulary	6-9%	1112.RWL.b1 Use context (e.g., the overall meaning of a sentence, paragraph, or text; a word's position in a sentence) as a clue to the meaning of a word or phrase	SR	3-4	3-4
		1112.RWL.c3 Develop and explain ideas for why authors made specific word choices within text	SR		
Writing	36-40%	1112.WI.b2 Create an organizational structure for writing that groups information logically (e.g., cause/effect, compare/contrast, descriptions and examples) to support paragraph focus	SR	2-4	2-4
		1112.WI.b4 Select the facts, extended definitions, concrete details, quotations, or other information and examples that are most relevant to the focus and appropriate for the audience	SR		
		1112.WP.f1 Produce a clear, coherent, permanent product that is appropriate to the specific task, purpose (to persuade), and audience	MSR, CR	7	15
Total	**100%**			**39 Total Items**	**47 Total Points**

** Percentages are approximate with the total equaling 100%*
*** CCCs require a multipart item to assess.*
****MSR- Multi-Select Response: for MSAA, this item type is a two-part Selected Response with a dependency between items, similar to an Evidence-Based Selected Response (EBSR). Students are not allowed to reference the previous item because the answer to the first item in the pair is included in the directions/stem of the second item.*

ELA Content Category	HS
Reading Literary	17-26%
Reading Informational	32-36%
Reading Vocabulary	6-9%
Writing	36-40%

20-21 MSAA Mathematics Operational Blueprint

- * Standards with operational CR items in 2019
- ** Standards with operational CR items beginning in 2020 and 2021

Grade 3 Targets by Standard

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range
Operations and Algebraic Thinking	28-32%	3.NO.2d3 Solve multiplication problems with neither number greater than 5	SR	10 -11
		3.NO.2e1 Solve or solve and check one- or two-step word problems requiring addition, subtraction, or multiplication with answers up to 100		
		3.PRF.2d1 Identify multiplication patterns in a real word setting		
Number and Operations Base Ten	17-23%	3.NO.1j3 Use place value to round to the nearest 10 or 100	SR CR	7
		3.NO.2c1** Solve multi-step addition and subtraction problems up to 100		
Number and Operations Fractions	17-23%	3.NO.1l3 Identify the fraction that matches the representation (rectangles and circles; halves, fourths, thirds, and eighths)	SR	7
		3.SE.1g1 Use =, <, or > to compare 2 fractions with the same numerator or denominator		
Measurement and Data	17-23%	3.DPS.1g1* Collect data; organize into picture or bar graph	SR CR	7
		3.ME.1d2 Measure area of rectilinear figures by counting squares		
Geometry	9-11%	3.GM.1i1 Partition rectangles into equal parts with equal area	SR	3 -4
Total	**100%**			**35**

Grade 4 Targets by Standard

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range
Operations and Algebraic Thinking	28-32%	4.NO.2d7 Determine how many objects go into each group when given the total number of objects and groups where the number in each group or number of groups is not > 10	SR	10-11
		4.PRF.1e3 Solve multiplicative comparisons with an unknown using up to 2-digit numbers with information presented in a graph or word problem (e.g., an orange hat cost $3. A purple hat cost 2 times as much. How much does the purple hat cost? [3 x 2 = p])		
		4.NO.2e2 Solve or solve and check one or two step word problems requiring addition, subtraction, or multiplication with answers up to 100		
Numberand Operations Base	9-11%	4.NO.1j5 Use place value to round to any place (i.e., ones, tens, hundreds, thousands)	SR	3-4
Number and Operations Fractions	28-32%	4.NO.1m1 Determine equivalent fractions	SR	10-11
		4.NO.1n2 Compare up to 2 given fractions that have different denominators		
		4.SE.1g2 Use =, <, or > to compare 2 fractions (fractions with a denominator or 10 or less)		
Measurem ent and Data	17-23%	4.ME.1g2 Solve word problems using perimeter and area where changes occur to the dimensions of a rectilinear figure	SR CR	7
		4.DPS.1g3* Collect data; organize in graph (e.g. picture graph, line plot, bar graph)		
Geometry	9-11%	4.GM.1h2* Classify two-dimensional shapes based on attributes (# of angles)	SR CR	3-4
Total	100%			35

Grade 5 Targets bt Standard

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range
Operations and Algebraic Thinking	9-11%	5.PRF.2b1 Generate or select a comparison between two graphs from a similar situation	SR	3-4
Number and Operations Base Ten	34-40%	5.NO.1b1 Read, write, or select a decimal to the hundredths place	SR CR	14
		5.NO.1b4 Round decimals to the next whole number		
		5.NO.2c1 Solve one-step problems using decimals		
		5.NO.2a5** Solve word problems that require multiplication or division		
Number and Operations Fractions	17-23%	5.NO.2c2 Solve word problems involving the addition, subtraction, multiplication, or division of fractions	SR	7
		5.PRF.1a1 Determine whether the product will increase or decrease based on the multiplier		
Measurement and Data	17-23%	5.ME.1b2 Convert standard measurements of length	SR	7
		5.ME.2a1 Use a calculator to solve one-step problems involving conversions of standard measurement units of area, volume, time, mass in the same system		
Geometry	9-11%	5.GM.1c3* Use order pairs to graph given points	SR CR	3-4
Total	100%			35

Grade 6 Targets by Standard

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range
Ratio and Proportions	28-32%	6.PRF.1c1 Describe the ratio relationship between two quantities for a given situation	SR	10-11
		6.ME.2a2 Solve one-step real world measurement problems involving unit rates with ratios of whole numbers when given the unit rate (3 inches of snow falls per hour, how much in 6 hours?)		
		6.NO.1f1 Find a percent of a quantity as rate per 100		
Expressions and Equations	17-23%	6.PRF.1d1 Solve real world single-step linear equations	SR	7
		6.NO.2a6 Solve problems or word problems using up to three-digit numbers and any of the four operations		
The Number System	28-32%	6.NO.2c3 Solve one-step, addition, subtraction, multiplication, or division problems with	SR CR	10-11
		6.NO.1d4** Select the appropriate meaning of a negative number in a real world situation		
		6.NO.1d2* Locate positive and negative numbers on a number line		
Statistics and Probability	9-11%	6.DPS.1d3 Select the statement that matches mean, mode, and spread of data for 1 measure of central tendency for a given data set	SR	3-4
Geometry	9-11%	6.GM.1d1 Find area of quadrilaterals	SR	3-4
Total	**100%**			**35**

Grade 7 Targets by Standard

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range
Ratio and Proportions	34-40%	7.NO.2f1** Identify the proportional relationship between two quantities (use rules or symbols to show quantitative relationships)	SR CR	14
		7.NO.2f2 Determine if two quantities are in a proportional relationship using a table of equivalent ratios or points graphed on a coordinate plane		
		7.NO.2f6 Solve word problems involving ratios		
		7.PRF.1f1 Use proportional relationships to solve multistep percent problems in real world situations		
Expressions and Equations	9-11%	7.PRF.1g2 Use variables to represent quantities in a real-world or mathematical problem, and construct simple equations and inequalities to solve problems by reasoning about the quantities	SR	3-4
The Number System	17-23%	7.NO.2i1 Solve multiplication problems with positive/negative numbers	SR	7
		7.NO.2i2 Solve division problems with positive/negative numbers		
Statistics and Probability	9-11%	7.DPS.1k1* Analyze graphs to determine or select appropriate comparative inferences about two samples or populations	SR CR	3-4
Geometry	17-23%	7.ME.2d1 Apply formula to measure area and circumference of circles	SR	7
		7.GM.1h2 Find the surface area of three-dimensional figures using nets of rectangles or triangles		
Total	100%			35

Grade 8 Targets by Standard

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range
Functions	17-23%	8.PRF.2e2** Identify the rate of change (slope) and initial value (*y*-intercept) from graphs	SR	7
		8.PRF.1f2 Describe or select the relationship between the two quantities given a line graph of a situation		
Expressions and Equations	17-23%	8.PRF.1e2 Represent proportional relationships on a line graph	SR	7
		8.PRF.1g3 Solve linear equations with 1 variable		
The Number System	9-11%	8.NO.1k3* Use approximations of irrational numbers to locate them on a number line	SR CR	3-4
Statistics and Probability	17-23%	8.DPS.1h1* Graph bivariate data using scatter plots and identify possible associations between the variables	SR CR	7
		8.DPS.1k2 Analyze displays of bivariate data to develop or select appropriate claims about those data		
Geometry	28-32%	8.ME.1e1 Describe the changes in surface area, area, and volume when the figure is changed in some way (e.g., scale drawings)	SR	10-11
		8.GM.1g1 Recognize congruent and similar figures		
		8.ME.2d2 Apply the formula to find the volume of 3-dimensional shapes (i.e., cubes, spheres, and cylinders)		
Total	100%			35

High School Targets by Standard

Content Category	Weight	Core Content Connector	Item Type	2021 Item Range
Algebra And Functions	47-52%	H.PRF.2b1** Translate a real-world problem into a one-variable linear equation	SR CR	17-18
		H.PRF.2b2 Solve equations with one or two variables using equations or graphs		
		H.ME.1b2 Solve a linear equation to find a missing attribute given the area, surface area, or volume and the other attribute		
		H.PRF.1c1 Select the appropriate graphical representation of a linear model based on real world events		
		H.PRF.2c1 Make predictions based on a given model (for example, a weather model, data for athletes over years)		
Number and Quantity	17-23%	H.ME.1a2 Solve real world problems involving units of measurement	SR	7
		H.NO.1a1 Simplify expressions that include exponents		
Statistics and Probability	17-23%	H.DPS.1b1* Complete a graph given the data, using dot plots, histograms, or box plots	SR CR	7
		H.DPS.1c1 Use descriptive stats, range, median, mode, mean, outliers/gaps, to describe data set		
Geometry	9-11%	H.GM.1b1 Use definitions to demonstrate congruency and similarity in figures	SR	3-4
Total	**100%**			**35**

APPENDIX D

PANELISTS AND COMMITTEE MEMBERS

Table D-1. MSAA 2023 Item Content and Bias Review Meeting & 2022 Passage Review Panelists by Grade and Content Area

ELA Content Grades 3–6	
Name	**State**
Anntonelli Pola	AS
Nellisa Delgado	DC
Abigail Trask	ME
Racheal Ahearn	ME
Omar Tabb	AZ
Mary Ashes	SD
Sonya Hebert	SD
Rhonda Gross	CNMI
Edward Desiderio	AS

ELA Content Grades 7, 8, HS	
Name	**State**
Christina Marino	DC
Gaye McNeil	MT
Mary Fried	SD
Tiffany Christman	SD
Deborah Karppala	AZ
Tracy Lynn Del Rosario	CNMI
Helene Cruz	Guam
Tagiilima Uikirifi	AS
Gretchen Lehmann	BIE

ELA Bias All Grades	
Name	**State**
Jesssica Fatina	AZ
Richard Kisa Masiwemai	CNMI
Christine Hernandez	Guam
Alaina Fisk	AZ
Deanna Viola-Bennett	ME
Senerita Kaleopa-Palauni	AS
Monica Waltman	SD
Carolyn Boughton	AZ
Thomas Vu	DC
Melanie Bowlus	ME

ELA Content Grades 3–5 Passage Review	
Name	**State**
Shelby Thibodeau	ME
Teneice Kirby	TN
Elizabeth White	TN
Joanna Moses	AZ
Nicole Poole	ME
Grace Diekman	MT

ELA Content Grades 8 and HS Passage Review	
Name	**State**
Andrew Sumner	ME
Morgan Brewer	ME
Hannah Pack	TN
Victoria Doiron	AZ
Danielle Barnes	AZ
Andrew Alpers	TN
Bonnie Iannaccone	TN

Mathematics Content Grades 3–6	
Name	**State**
Amy Dawson	AZ
Mark Bennett	ME
Kailie Balcom	ME
Kelley Wallace	ME
Lizelle Amirez	CNMI
Ashley Wilder	AZ
Darla Stone	MT
Lan Moi	AS
Krystal Butler	DC

Mathematics Content Grades 7, 8, HS	
Name	**State**
Brandon Bernard	AZ
Gloria Tuigamala	AS
Becky Erickson	SD
Lynette Busch	SD
Andrea Bozzino	ME
Priscila Gomez	ME
Rebecca Dominguez	AZ
Atalina Coffin	AS

Mathematics Bias All Grades	
Name	**State**
Kimberly Aikins	AZ
Jessica L. Wasisang-Mendez	CNMI
Emma Haberman	SD
Lauren Thompson	DC
Misty Favreau	ME
Rucelio Vilar	BIE
Roberto Cruz	AZ
Cassandra La'apui	AS

ELA Content Grades 6–7 Passage Review	
Name	**State**
Kelley Wallace	ME
Tina Hicks	ME
Roch Turner	MT
Jeremy Hutson	TN
Loretta Donovan	AZ
Amanda Gonzales	AZ
Jay Humphreys	TN
Audra Cole	ME
Doris Diras	AZ

Table D-2. MSAA 2023 Technical Advisory Committee Members

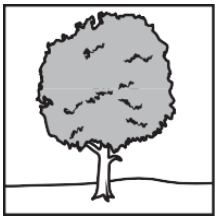
Name	Organization	Expertise
Derek Briggs	University of Colorado	• Assessment • Growth • Psychometrics
Chris Domaleski	Center for Assessment	• Accountability Systems • Psychometrics
Rachel Quenemoen	National Center on Educational Outcomes	• Students with Significant Cognitive Difficulties • NCSC Awareness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

APPENDIX E
TEACHER GUIDE—SAMPLE ITEMS

Sample Items 1 & 2		
Alignment	**Core Content Connector (CCC): 3.RI.k5** Determine the main idea of a text; recount the key details and explain how they support the main idea.	
Learning Targets	**Instructional Strategies**	**Scaffolds and Supports**
I can determine the topic of an informational text presented in diverse media. I can identify a supporting detail of the topic in a text. I can identify a supporting detail in diverse media that supports the topic in the medium.	**Graphic Organizer** • List the topic of a text or multimedia and note events and/or details that support the topic. • Use a System of Least Prompts when selecting a supporting detail. **Topic Board/Display** • Identify pictures that represent the topic(s) of a given text. • Include illustrations or sentences from the text; include events and details that support the topic in a topic board/display or graphic organizer. **Interactive Story Reading** • Choose and pre-read a text prior to instruction. • Read the text aloud to students, stopping at predetermined points. • At each stopping point, ask students to share their thoughts and respond to text. **Group Think** Tell the students what the topic is prior to reading a text or watching multimedia. After reading the text, ask the students to identify sentences that tell you the topic and supporting details about the topic. **Think Aloud** Model the thought processes that occur while reading the text. This may include asking questions while reading the text, identifying important details, identifying the topic, and identifying the main idea.	• Pictures, objects, or tactile representations to illustrate the topic, events, or details • Sentence strips that reflect supporting details about the topic • Videos or storyboards/ cards of the story for visual supports. • Technology (e.g., interactive whiteboard, informational texts read by the computer that highlights text)

Item 1*

What is the main idea in this passage?



A. The Sun helps trees to grow big and tall.

B. People can guess a riddle and win a prize.

C. The Lantern Festival is important to families.

Would you like to read this question again, yes or no?

*__Please note:__ passage may be accessed in the sample items PDF and Directions for Test Administration.

APPENDIX F
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	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
Language	04				
Arts	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
	○ 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: ○ + = Correct Response ○ A,B,C,D = Incorrect Response ○ Z = No Response ○ X = Item Excluded from Student's form • Writing Prompt: ○ 0,1,2 = Response Score ○ Z (blank) ,F (Foreign Language) ,P (Copy of Prompt) ,N (No Score) ,O (Off Topic) Non-Scorable Codes ○ 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 ○ 0,1,2 = Response Score ○ 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 G
GUIDE FOR SCORE REPORT INTERPRETATION

2023 Guide for Score Report Interpretation

Multi-State Alternate Assessment (MSAA). 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 11. 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.

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.

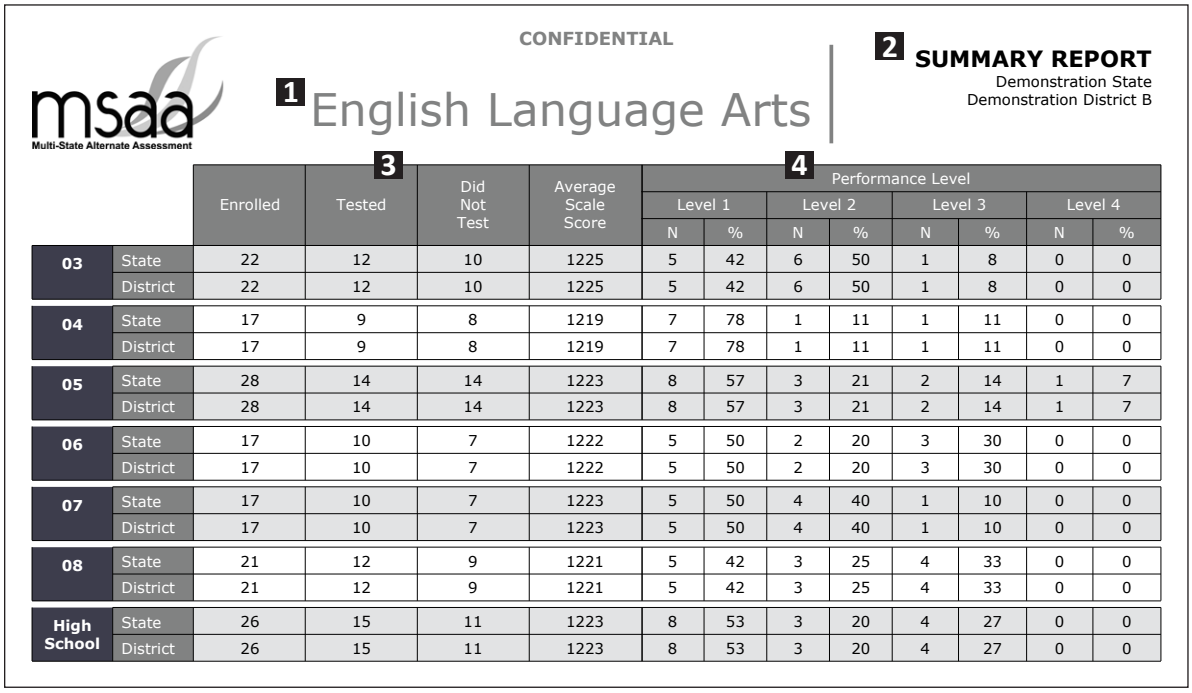
All submitted tests receive a participation status, regardless of the number of item responses.

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

msaa
Multi-State Alternate Assessment

1 English Language Arts

2 SUMMARY REPORT
Demonstration State
Demonstration District B

		3 Enrolled	Tested	Did Not Test	Average Scale Score	4 Performance Level							
						Level 1		Level 2		Level 3		Level 4	
						N	%	N	%	N	%	N	%
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

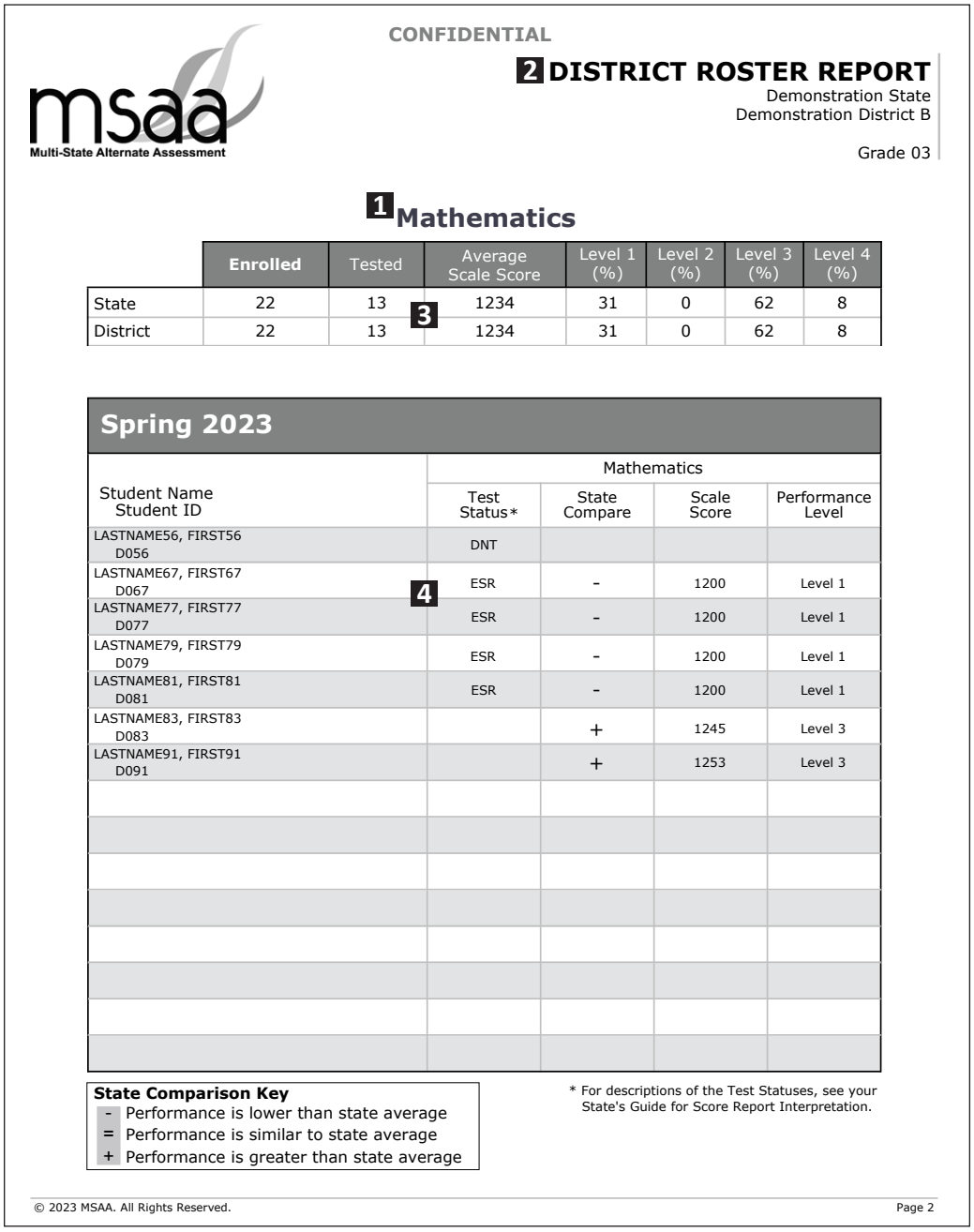
The DSR contains the following features, highlighted above:

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

msaa
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.

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

CONFIDENTIAL

msaa
Multi-State Alternate Assessment

1 English Language Arts

2 SUMMARY REPORT
Demonstration State
Demonstration District B
Demonstration School 4

		3 Enrolled	Tested	Did Not Test	Average Scale Score	4 Performance Level							
						Level 1		Level 2		Level 3		Level 4	
						N	%	N	%	N	%	N	%
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

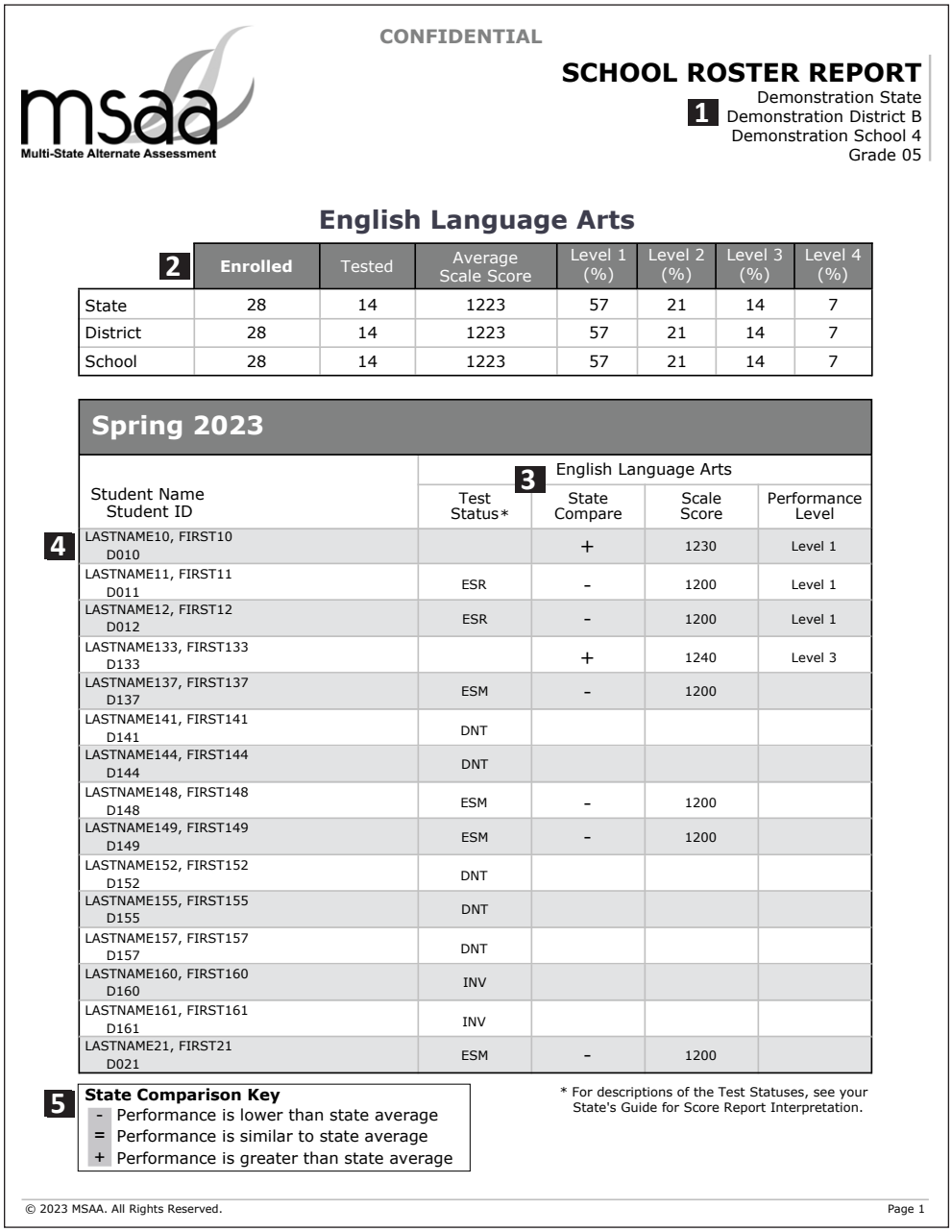
The SSR contains the following features, highlighted above:

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



CONFIDENTIAL

msaa
Multi-State Alternate Assessment

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

Student Name / Student ID	3 English Language Arts: 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.

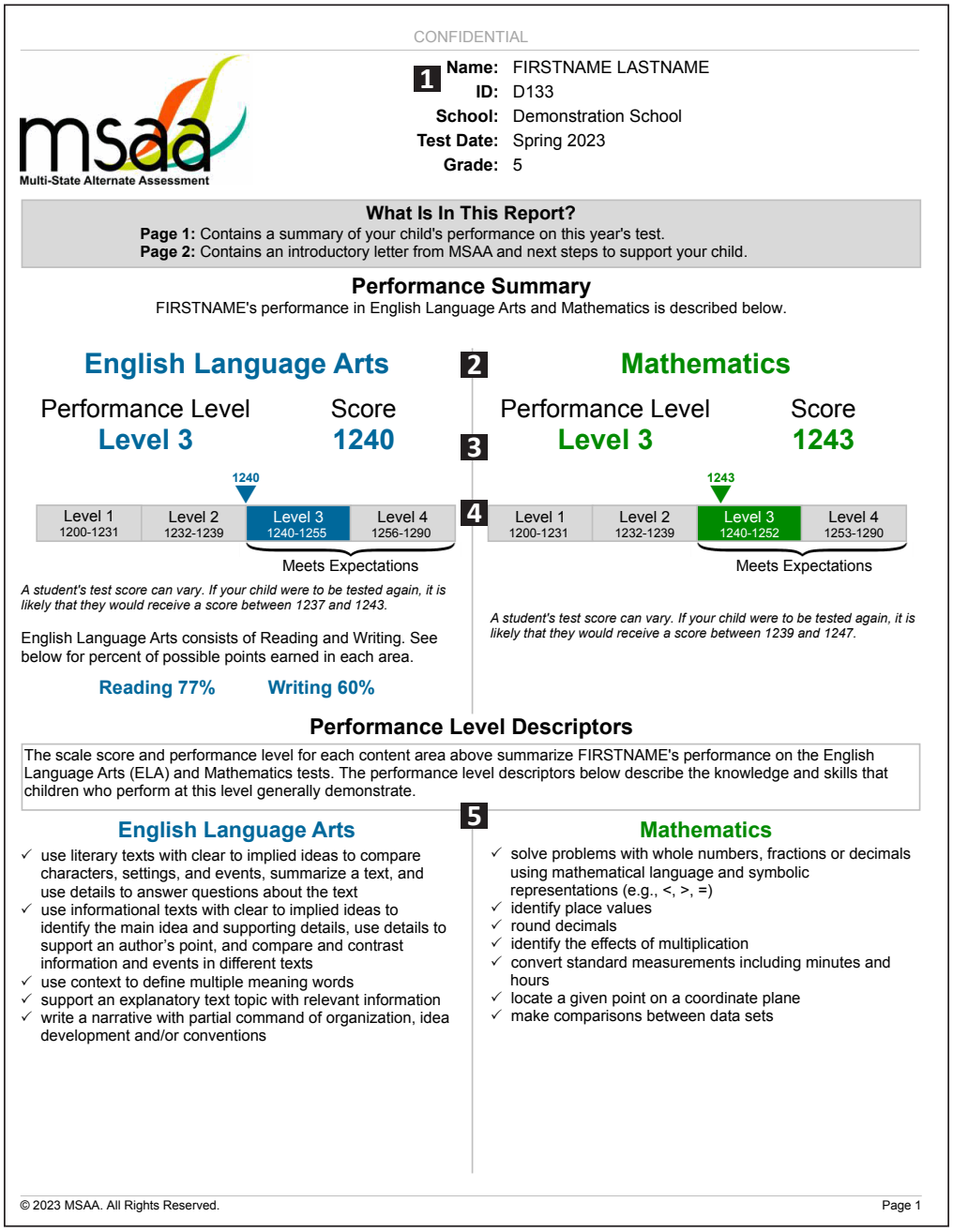
The SRR contains the following features, highlighted above:

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.

English Language Arts

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

2
3
4

1240

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

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.

[Page is intentionally blank]

Appendix A: Writing Scoring Rubrics

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	
Organization – 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 **two** opposing conditions • a body that includes: o **one** activity **common to both** conditions o **one** activity related to **each** of the two opposing conditions • a conclusion that states the **two** opposing conditions	The essay includes at a minimum: • an introduction that presents the topic • a body that includes: o **one** activity **common to both** conditions o **one** activity related to **one of the two** opposing conditions • a conclusion that states 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: • **one** activity related to **both** conditions **with** a relevant detail • **one** activity related to **each** of the two opposing conditions, **each with** relevant details	The essay includes at a minimum: • **two** activities **each with** a relevant detail	The essay includes at a minimum: • **one** activity **OR** • **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., 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	

Grade 7 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 with an effect related directly to a cause (e.g., cause/effect).	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 **and** its effect	The essay includes at a minimum: • an introduction that states the topic/cause • a body that includes an effect that **may not** relate to the provided cause • a conclusion that states a cause **or** the effect	The essay includes at a minimum: • **some** evidence related to the specified topic (i.e., introduction, cause/effect relationship, **or** conclusion)	**0** • **no** evidence of organization	**5** • 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 effect	The essay includes at a minimum: • **one** effect with **no** relevant detail	The essay includes at a minimum: • **one** idea related to the topic	**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 7 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 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 **and** its effects • a body that includes **two** effects **and** refers them to the cause • a conclusion that states the essay is about a cause **and** its effects	The essay includes at a minimum: • an introduction that presents a topic • a body that includes **one** effect **and** refers it to the cause • a conclusion that states the topic	The essay includes at a minimum: • **some** evidence related to the specified topic (i.e., introduction, on-topic cause/effect 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: • **two** effects, **each** with a relevant detail • transitional words to connect the cause to **each** of the **two** effects	The essay includes at a minimum: • **one** effect **with** a relevant detail • transitional word to **connect one** cause/effect relationship	The essay includes at a minimum: • **one** detail that describes the cause **or** effect **OR** • **one** transition word	**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	

Grade 8 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 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)	**0** • **no** evidence of organization	**5** • 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	**0** • **no** evidence of idea development	**5** • 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., "**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	

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	

[Page is intentionally blank]

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

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 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	**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	**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	**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
	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 • use context to identify the meaning of multiple-meaning words	• 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	
	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 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	**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	**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	**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 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

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 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	**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	**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	**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
	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*	
	• 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	• 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	
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	**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	**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	**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

*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="4">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="4">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

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 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	**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	**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	**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
	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*	
	• solve multiplication problems with positive/ negative whole numbers • interpret graphs to qualitatively contrast data sets	• solve multiplication problems with positive/ negative whole numbers • evaluate variable expressions that represent word problems • interpret graphs to qualitatively contrast data sets	

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

Grade 8 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: • 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	**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	**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	**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
	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 congruent figures • use properties of similarity to identify similar figures • interpret data tables to identify the relationship between variables	• 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	

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

High School 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: • 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	**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	**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	**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
	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 the linear representation of a provided real-world situation • use an equation or a linear graphical representation to solve a word problem	• 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	

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

[Page is intentionally blank]

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

High School Science Performance Level Descriptors

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

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

High School Science Performance Level Descriptors

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

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.

Appendix D: Individual Student Report Samples

CONFIDENTIAL

msaa
Multi-State Alternate Assessment

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

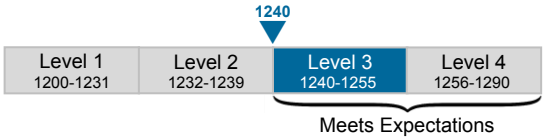
What Is In This Report?
Page 1: Contains a summary of your child's performance on this year's test.
Page 2: Contains an introductory letter from MSAA and next steps to support your child.

Performance Summary

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

English Language Arts

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



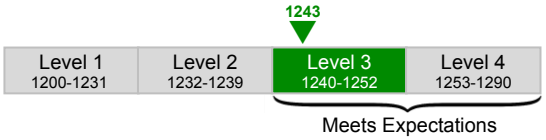
A student's test score can vary. If your child were to be tested again, it is likely that they would receive a score between 1237 and 1243.

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**



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

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?

Multi-State Alternate Assessment

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.

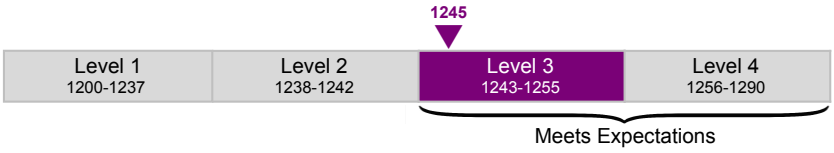
Performance Summary

FIRSTNAME's performance in Science is described below.

Science

Performance Level
Level 3

Score
1245



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?


msaa
Multi-State Alternate Assessment

APPENDIX H

DIFFERENTIAL ITEM FUNCTIONING RESULTS

Table H-1. DIF— ELA—Dichotomous

Grade	Group		Number of Items	Number "Low"			Number "High"		
	Reference	Focal		Total	Favoring		Total	Favoring	
					Reference	Focal		Reference	Focal
3	Male	Female	70	6	4	2	1	0	1
	Non-EconDis	EconDis	70	6	2	4	1	0	1
	White	Black or African American	70	2	1	1	1	0	1
		Hispanic or Latino	70	2	0	2	0	0	0
4	Male	Female	73	4	2	2	2	2	0
	Non-EconDis	EconDis	73	5	4	1	0	0	0
	White	Black or African American	73	7	4	3	1	1	0
		Hispanic or Latino	73	4	1	3	1	1	0
5	Male	Female	72	8	6	2	1	1	0
	Non-EconDis	EconDis	72	4	3	1	0	0	0
	White	Black or African American	72	2	1	1	0	0	0
		Hispanic or Latino	72	4	1	3	1	0	1
6	Male	Female	64	5	4	1	1	1	0
	Non-EconDis	EconDis	64	1	1	0	1	1	0
	White	Black or African American	64	8	3	5	1	1	0
		Hispanic or Latino	64	3	0	3	0	0	0
7	Male	Female	64	5	1	4	0	0	0
	Non-EconDis	EconDis	64	5	2	3	0	0	0
	White	Black or African American	64	4	1	3	0	0	0
		Hispanic or Latino	64	2	0	2	0	0	0
8	Male	Female	64	5	3	2	0	0	0
	Non-EconDis	EconDis	64	1	0	1	0	0	0
	White	Black or African American	64	8	3	5	0	0	0
		Hispanic or Latino	64	7	2	5	1	0	1
HS	Male	Female	73	4	3	1	0	0	0
	Non-EconDis	EconDis	73	2	1	1	0	0	0
	White	Black or African American	73	0	0	0	0	0	0
		Hispanic or Latino	73	4	3	1	0	0	0

Table H-2. DIF—ELA / Writing Prompt—Polytomous

Grade	Group		Number of Items	Number "Low"			Number "High"		
	Reference	Focal		Total	Favoring		Total	Favoring	
					Reference	Focal		Reference	Focal
3	Male	Female	6	0	0	0	0	0	0
	Non-EconDis	EconDis	6	0	0	0	0	0	0
	White	Black or African American	6	2	2	0	0	0	0
		Hispanic or Latino	6	1	1	0	0	0	0
4	Male	Female	6	0	0	0	0	0	0
	Non-EconDis	EconDis	6	0	0	0	0	0	0
	White	Black or African American	6	1	1	0	0	0	0
		Hispanic or Latino	6	1	1	0	0	0	0
5	Male	Female	6	1	0	1	0	0	0
	Non-EconDis	EconDis	6	1	0	1	0	0	0
	White	Black or African American	6	1	1	0	0	0	0
		Hispanic or Latino	6	0	0	0	1	1	0
6	Male	Female	6	2	0	2	0	0	0
	Non-EconDis	EconDis	6	2	2	0	0	0	0
	White	Black or African American	6	2	1	1	0	0	0
		Hispanic or Latino	6	2	2	0	0	0	0
7	Male	Female	6	0	0	0	0	0	0
	Non-EconDis	EconDis	6	1	1	0	0	0	0
	White	Black or African American	6	0	0	0	0	0	0
		Hispanic or Latino	6	0	0	0	0	0	0
8	Male	Female	6	1	0	1	0	0	0
	Non-EconDis	EconDis	6	0	0	0	0	0	0
	White	Black or African American	6	0	0	0	0	0	0
		Hispanic or Latino	6	0	0	0	0	0	0
HS	Male	Female	6	1	0	1	0	0	0
	Non-EconDis	EconDis	6	1	1	0	0	0	0
	White	Black or African American	6	1	0	1	0	0	0
		Hispanic or Latino	6	0	0	0	0	0	0

Table H-3. DIF— Mathematics—Dichotomous

Grade	Group		Number of Items	Number "Low"			Number "High"		
	Reference	Focal		Total	Favoring		Total	Favoring	
					Reference	Focal		Reference	Focal
3	Male	Female	69	4	2	2	1	1	0
	Non-EconDis	EconDis	69	2	1	1	0	0	0
	White	Black or African American	69	4	2	2	1	0	1
		Hispanic or Latino	69	3	1	2	0	0	0
4	Male	Female	67	3	2	1	0	0	0
	Non-EconDis	EconDis	67	5	1	4	1	0	1
	White	Black or African American	67	7	2	5	1	0	1
		Hispanic or Latino	67	2	1	1	0	0	0
5	Male	Female	70	4	2	2	0	0	0
	Non-EconDis	EconDis	70	2	0	2	0	0	0
	White	Black or African American	70	6	3	3	2	0	2
		Hispanic or Latino	70	8	6	2	0	0	0
6	Male	Female	68	4	2	2	0	0	0
	Non-EconDis	EconDis	68	5	1	4	0	0	0
	White	Black or African American	68	4	4	0	1	0	1
		Hispanic or Latino	68	3	2	1	0	0	0
7	Male	Female	71	2	0	2	0	0	0
	Non-EconDis	EconDis	71	5	2	3	0	0	0
	White	Black or African American	71	4	2	2	1	1	0
		Hispanic or Latino	71	5	1	4	0	0	0
8	Male	Female	70	5	2	3	0	0	0
	Non-EconDis	EconDis	70	1	0	1	0	0	0
	White	Black or African American	70	5	3	2	0	0	0
		Hispanic or Latino	70	4	1	3	0	0	0
HS	Male	Female	68	3	2	1	0	0	0
	Non-EconDis	EconDis	68	2	1	1	0	0	0
	White	Black or African American	68	9	4	5	0	0	0
		Hispanic or Latino	68	3	3	0	1	0	1

APPENDIX I
ITEM RESPONSE THEORY PARAMETERS

Image I-1. Boxplot of Item Difficulty by Level for ELA—Grade 3

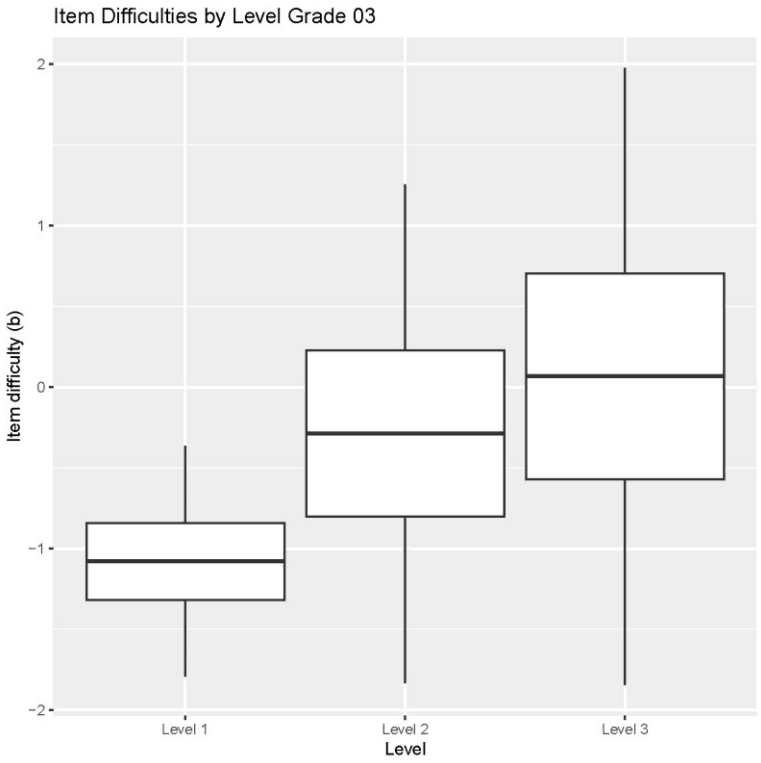


Image I-2. Boxplot of Item Difficulty by Level for ELA—Grade 4

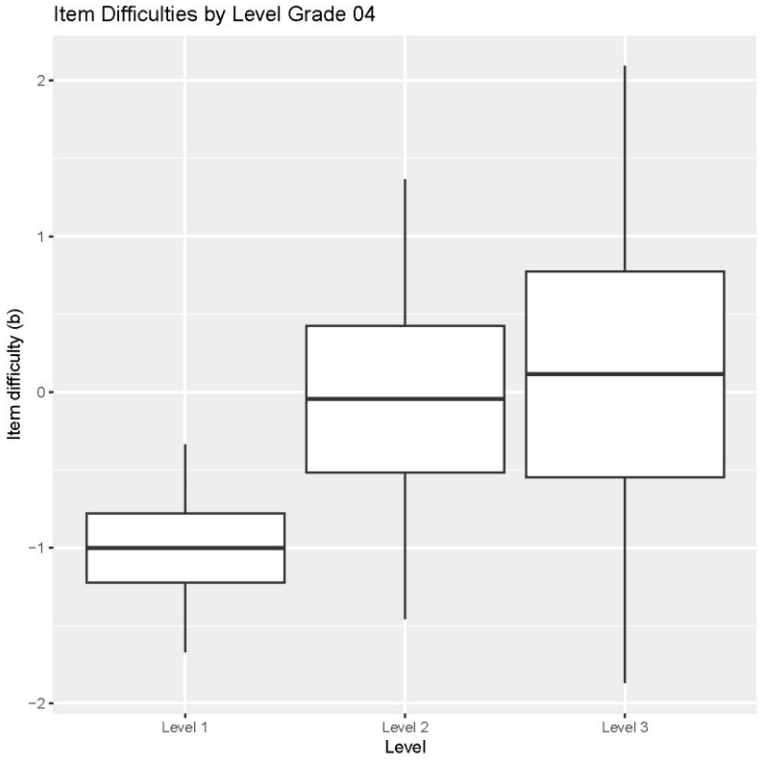


Image I-3. Boxplot of Item Difficulty by Level for ELA—Grade 5

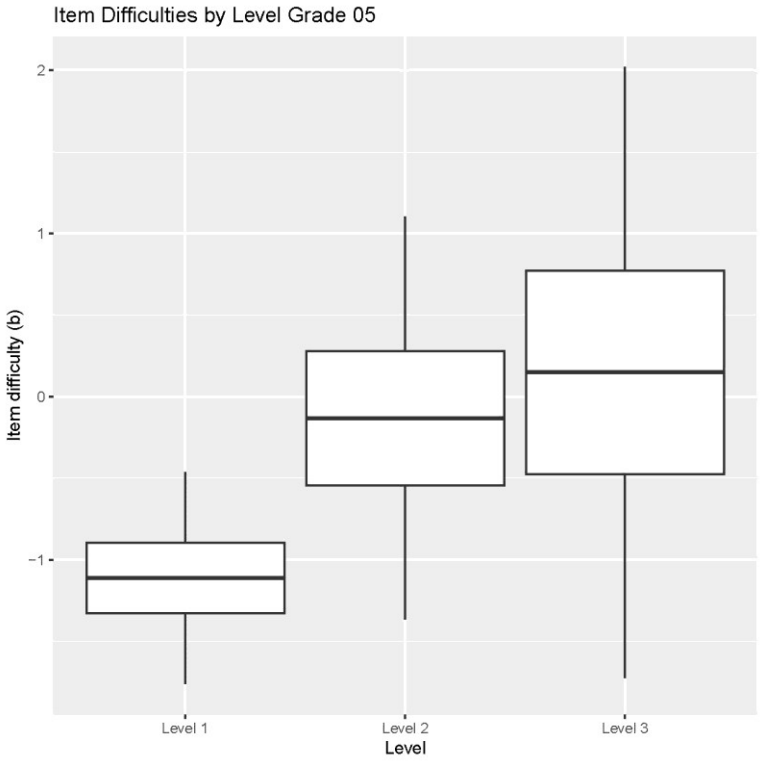


Image I-4. Boxplot of Item Difficulty by Level for ELA—Grade 6

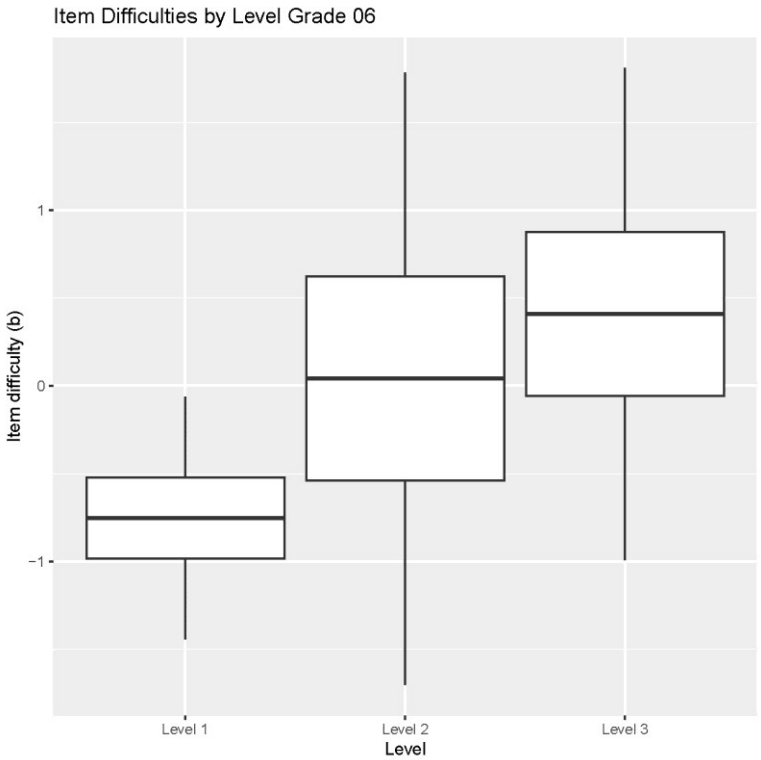


Image I-5. Boxplot of Item Difficulty by Level for ELA—Grade 7

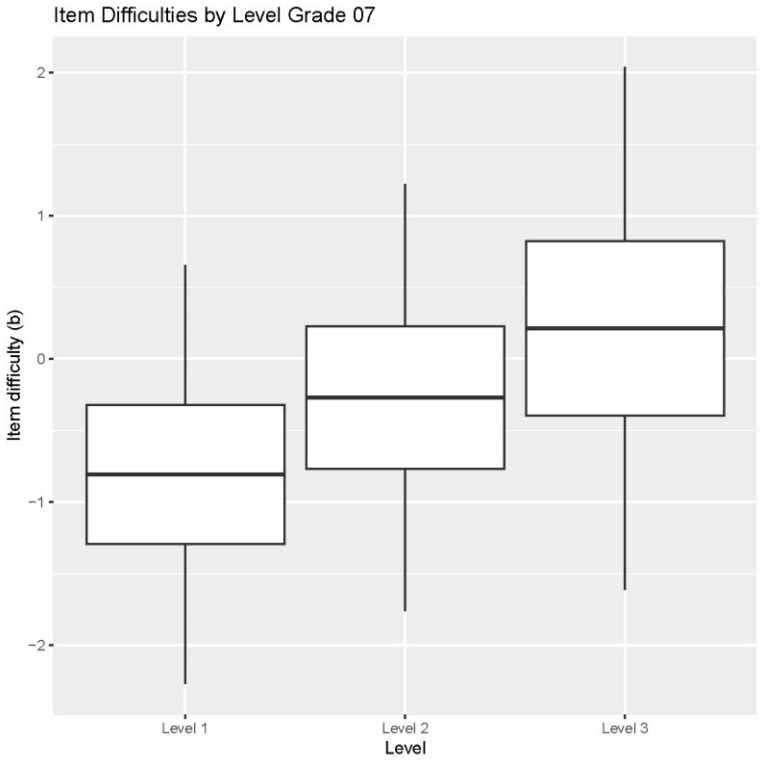


Image I-6. Boxplot of Item Difficulty by Level for ELA—Grade 8

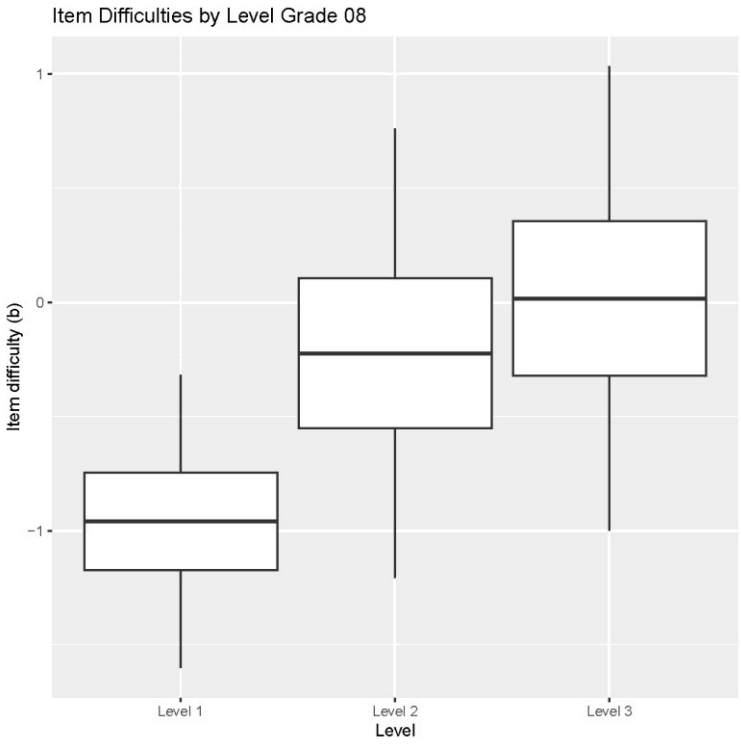


Image I-7. Boxplot of Item Difficulty by Level for ELA—Grade 11

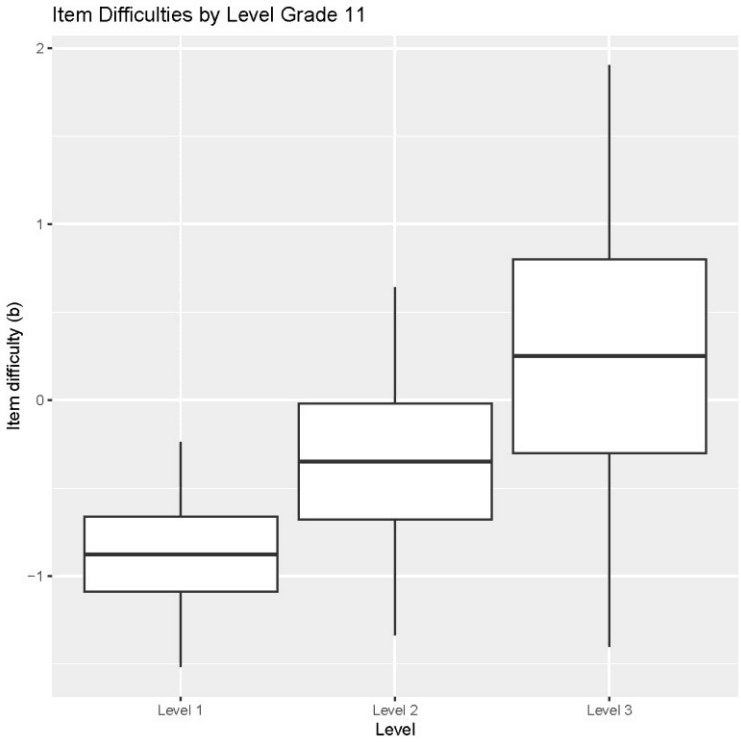


Image I-8. Boxplot of Item Difficulty for Mathematics—Grade 3

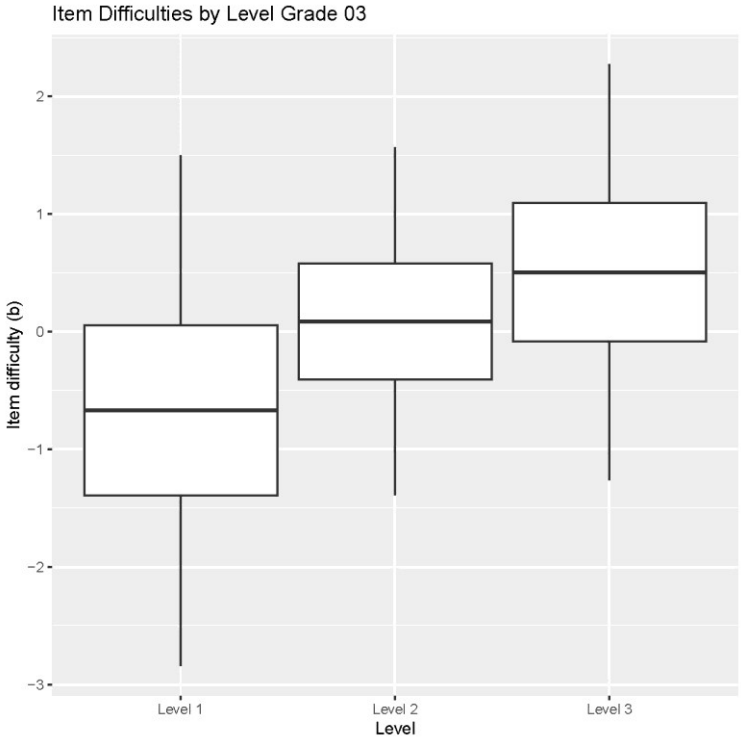


Image I-9. Boxplot of Item Difficulty for Mathematics—Grade 4

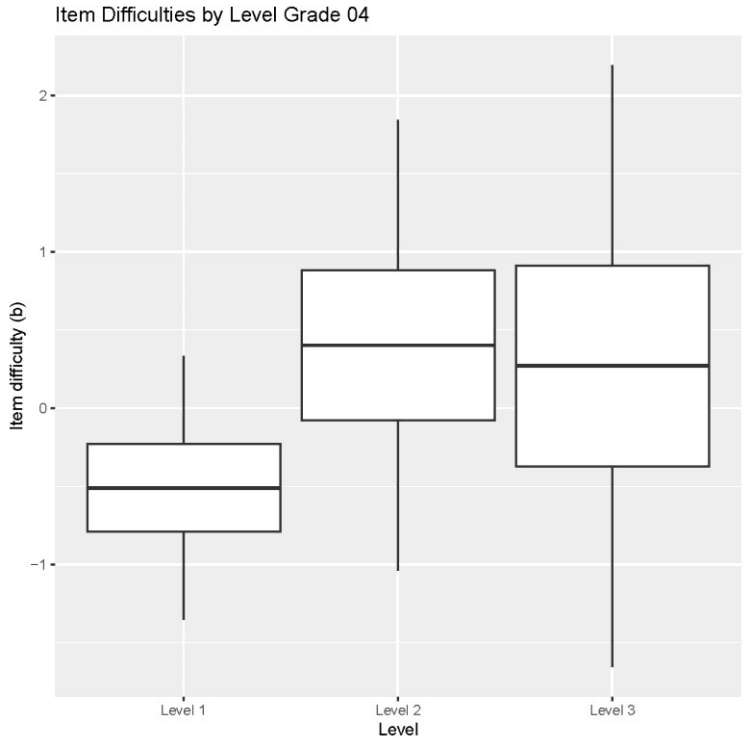


Image I-10. Boxplot of Item Difficulty for Mathematics—Grade 5

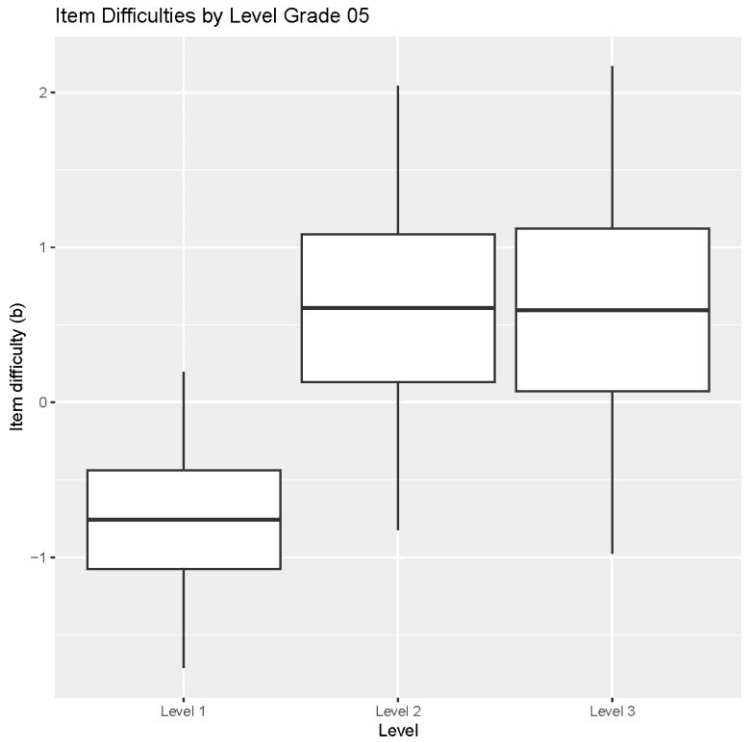


Image I-11. Boxplot of Item Difficulty for Mathematics—Grade 6

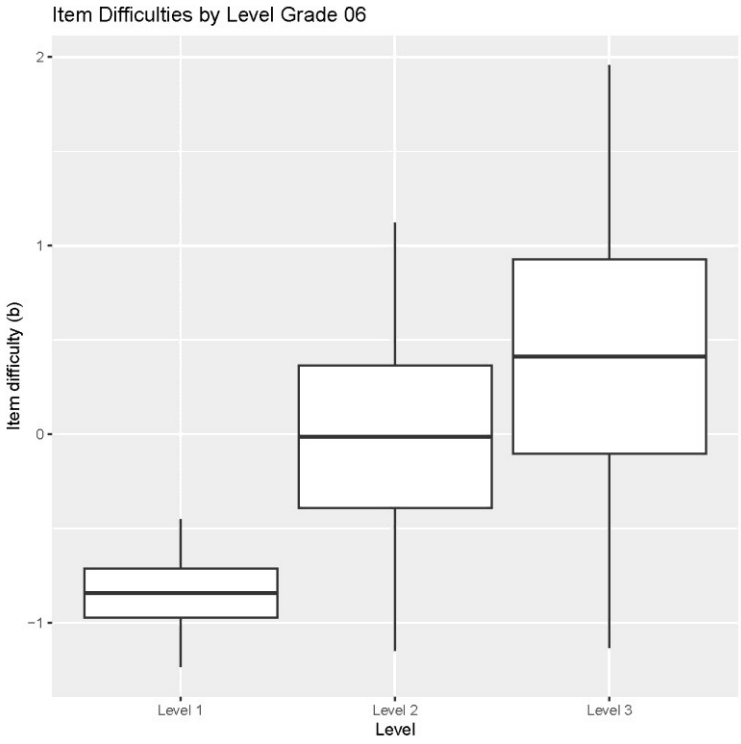


Image I-12. Boxplot of Item Difficulty for Mathematics—Grade 7

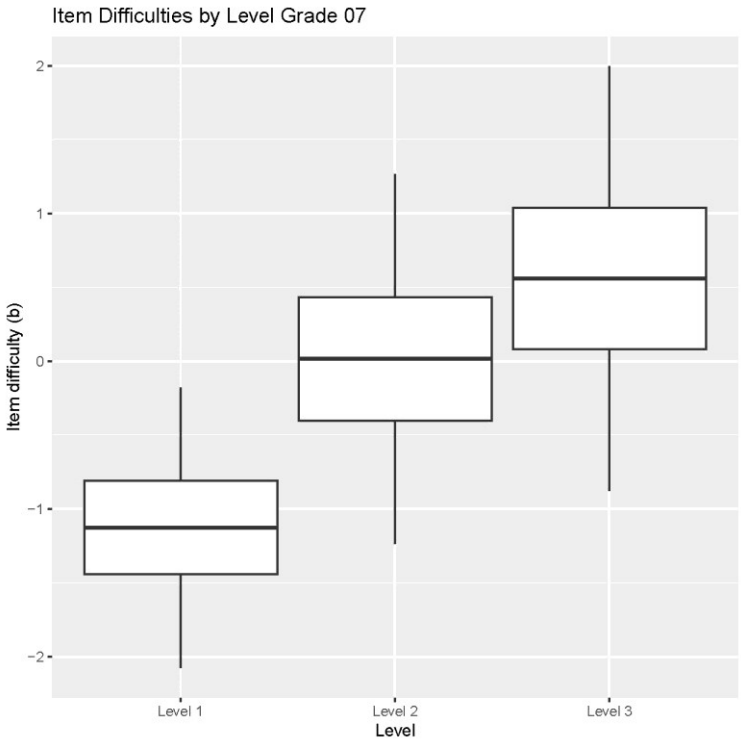


Image I-13. Boxplot of Item Difficulty for Mathematics—Grade 8

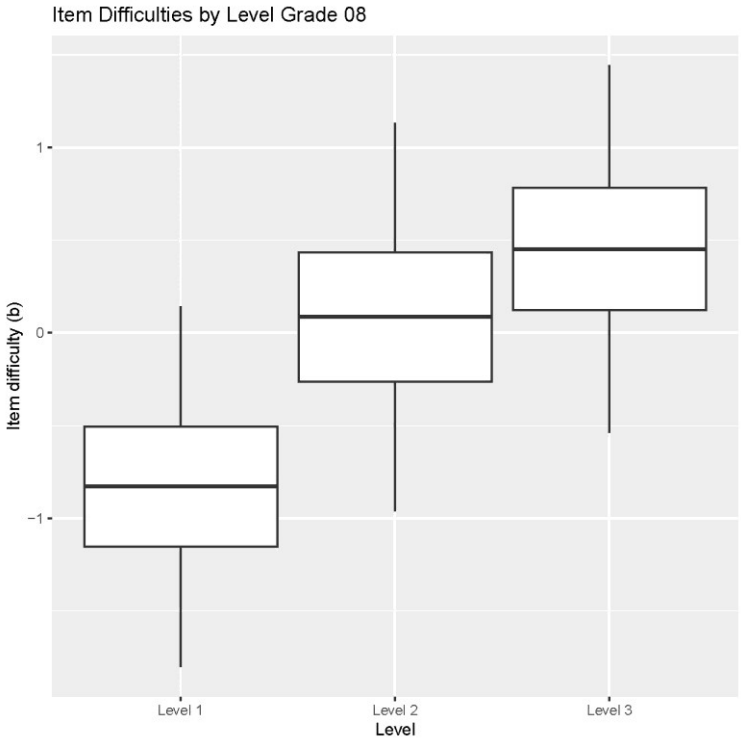


Image I-14. Boxplot of Item Difficulty for Mathematics—Grade 11

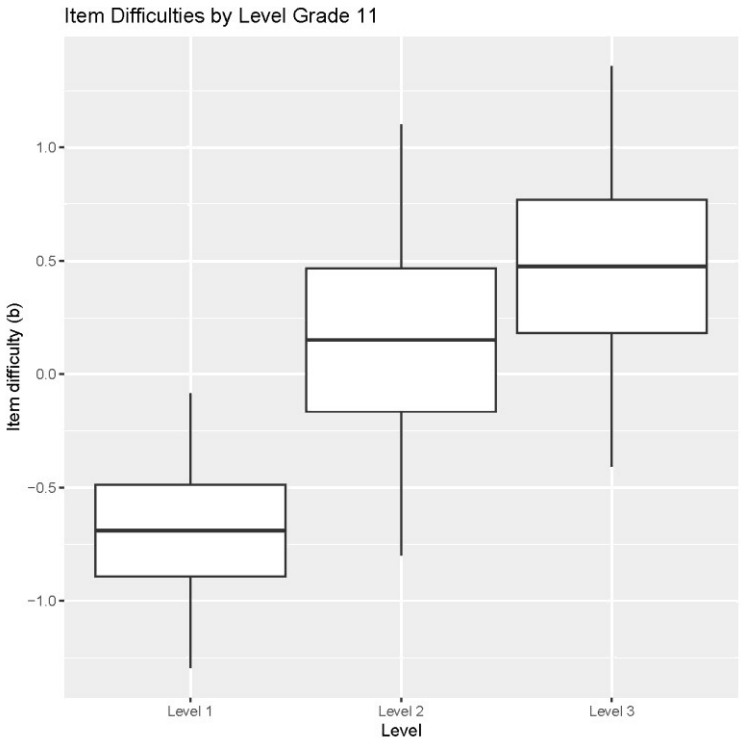


Table I-1. IRT Parameters for ELA Grade 3

IREF	a	SE (a)	b	SE (b)
113683A	0.53	0.08	-0.02	0.16
113682A	0.64	0.1	-0.38	0.18
114010A	1.07	0.1	-0.17	0.08
113746A	0.72	0.12	-0.42	0.08
113747A	1.07	0.13	-0.06	0.06
113681A	0.77	0.11	-0.54	0.19
114008A	1.69	0.28	-0.37	0.14
116203A	0.84	0.04	-0.29	0.04
116205A	0.63	0.04	-1.14	0.06
116204A	1.22	0.06	-0.58	0.03
120912A	0.95	0.12	-0.21	0.12
120914A	0.52	0.08	0.03	0.16
116202A	0.49	0.03	-0.12	0.06
120927A	0.43	0.09	-0.32	0.13
120926A	0.45	0.1	-0.25	0.12
121545A	0.96	0.09	-0.1	0.08
121731A	0.43	0.03	0.87	0.07
120922A	0.54	0.03	0.16	0.06
451521	0.52	0.11	0.92	0.25
451136	0.74	0.04	-0.37	0.04
451474	0.58	0.11	-0.91	0.16
451486	0.63	0.11	-0.42	0.09
121726A	0.65	0.11	-0.22	0.2
451172	1.2	0.09	0.14	0.05
451186	0.96	0.05	-0.57	0.03
451498	0.36	0.09	0.15	0.17
451534	1.04	0.13	-0.28	0.06
455391	1.26	0.1	-1.02	0.03
455415	1.16	0.1	-0.96	0.04
455427	1	0.13	-1.61	0.08
455439	1.01	0.09	-1.22	0.04
451148	0.44	0.03	-1.19	0.08
451160	0.54	0.03	-0.75	0.06
538072	1.45	0.11	-1.07	0.03
455403	0.83	0.12	-0.88	0.07
530009	1.06	0.07	-1.31	0.06
538024	0.94	0.13	-1.32	0.07
538036	1.45	0.11	-0.99	0.03
538060	1.06	0.14	-1.31	0.06
538048	1.35	0.15	-1.14	0.04
607594	1.13	0.09	-0.74	0.05
658866	0.48	0.08	-0.37	0.22
658842	0.89	0.1	0.34	0.08
659002	0.7	0.08	0.71	0.08
659083	0.93	0.13	0.06	0.09
659095	0.81	0.12	-0.24	0.11
659071	0.91	0.19	1.4	0.13
659143	0.68	0.1	0.44	0.11
658854	1.17	0.15	-0.29	0.11
658878	0.85	0.19	1.55	0.16
772760	1.11	0.09	-0.46	0.05
772753	0.89	0.09	-0.85	0.06
772774	1.13	0.13	0.05	0.07
675582	0.65	0.05	-0.24	0.07
675606	0.75	0.06	-0.57	0.06
772767	0.91	0.09	-0.85	0.06
780483	0.87	0.09	-0.71	0.07
786878	0.47	0.05	0.01	0.11
788711	0.8	0.07	-0.24	0.07
780468	1.17	0.12	-0.78	0.05
780504	1.06	0.11	-0.84	0.06
786871	0.67	0.08	-1.05	0.09
786892	0.78	0.1	0.52	0.1
786913	0.8	0.08	-0.72	0.07
780475	0.62	0.08	-1.29	0.13
780490	1	0.11	-1.03	0.07
780497	0.81	0.09	-1.07	0.09
786864	0.96	0.11	-1.22	0.08
787943	0.5	0.08	0.84	0.14
786899	0.69	0.06	0.54	0.1

Table I-2. IRT Parameters for ELA Grade 4

IREF	a	SE (a)	b	SE (b)
113095A	0.93	0.12	-0.29	0.07
113280A	0.68	0.04	-0.73	0.06
113283A	0.68	0.04	-0.65	0.06
113055A	1.47	0.15	-0.3	0.05
114056A	1.25	0.14	-0.31	0.05
114053A	1.23	0.14	-0.28	0.05
114054A	0.85	0.11	0.3	0.07
113281A	0.71	0.04	-0.26	0.04
116574A	0.69	0.08	-1.11	0.07
116576A	0.7	0.07	-1.37	0.07
121279A	1.15	0.09	-0.99	0.03
121550A	0.58	0.03	-0.61	0.05
113096A	1.53	0.15	-0.21	0.04
121426A	1.35	0.1	-0.89	0.03
121569A	1.26	0.14	-0.21	0.05
114055A	0.41	0.04	0.58	0.08
121551A	0.35	0.02	-0.43	0.08
451634	0.72	0.08	0.54	0.09
451646	0.57	0.09	0.4	0.13
451694	0.69	0.09	0.76	0.09
451913	0.38	0.03	0.34	0.09
449675	0.99	0.06	0.22	0.04
451598	0.55	0.08	-0.15	0.09
451867	0.49	0.04	-0.83	0.08
451574	0.71	0.09	-0.02	0.07
451586	0.64	0.11	0.92	0.16
451679	1.11	0.12	0.51	0.07
451895	0.44	0.03	0.05	0.06
451925	0.8	0.04	-0.32	0.04
122582A	0.32	0.08	-0.17	0.17
451550	1.02	0.11	-0.57	0.08
451562	1.07	0.13	-0.16	0.06
451663	0.72	0.08	0.39	0.1
451881	0.72	0.04	-0.28	0.04
528771	0.91	0.05	-0.97	0.04
615079	0.55	0.05	-0.46	0.09
615137	0.8	0.11	0.28	0.11
615172	1.48	0.1	-1.05	0.03
615065	0.9	0.07	-0.65	0.06
608509	0.62	0.03	-0.8	0.05
615093	0.77	0.15	0.89	0.13
615109	0.82	0.07	-0.86	0.07
615160	1.31	0.1	-1.09	0.03
615184	1.67	0.11	-1.01	0.03
659413	0.61	0.1	0.56	0.12
615196	1.16	0.1	-0.9	0.05
659377	0.82	0.08	-0.39	0.08
659401	0.91	0.11	0.62	0.08
677341	0.87	0.1	-1	0.09
677369	0.65	0.08	-1.14	0.13
677383	1.03	0.17	-0.35	0.14
677253	0.81	0.08	-0.06	0.08
677265	0.64	0.1	0.56	0.12
659365	0.46	0.08	0.38	0.16
659389	0.64	0.1	0.5	0.12
772934	0.97	0.11	-1.3	0.08
772941	1.37	0.13	-0.95	0.05
772969	1.23	0.2	-0.65	0.14
773092	0.51	0.05	-0.2	0.09
772948	1.15	0.13	-1.3	0.07
773099	0.52	0.08	0.73	0.13
677313	0.9	0.11	-1.12	0.1
677353	0.4	0.06	-1.02	0.18
677325	1.1	0.12	-0.89	0.07
772927	1.24	0.11	-0.92	0.05
772955	1.06	0.11	-1.12	0.07
773106	0.52	0.05	-0.53	0.09
773120	0.49	0.08	0.74	0.14
789700	0.83	0.07	-0.53	0.06
789714	0.76	0.08	0.12	0.07
789734	0.88	0.11	0.74	0.09
789723	0.94	0.08	-0.3	0.06
789730	1.69	0.17	-1.12	0.04

Table I-3. IRT Parameters for ELA Grade 5

IREF	a	SE (a)	b	SE (b)
114072A	0.91	0.12	-0.41	0.15
114331A	0.62	0.08	-0.24	0.12
114332A	0.84	0.11	-0.38	0.06
114073A	0.61	0.05	-0.26	0.07
114074A	0.26	0.04	0.63	0.17
117524A	0.31	0.03	-0.02	0.09
117523A	0.52	0.03	-0.33	0.05
117525A	0.51	0.03	-0.02	0.06
120909A	0.81	0.05	-0.4	0.04
120911A	0.72	0.05	-0.21	0.06
120913A	0.91	0.06	-0.06	0.05
121733A	0.52	0.1	0.07	0.11
121735A	0.63	0.1	-0.1	0.08
114075A	0.63	0.09	0.86	0.1
114329A	0.73	0.09	-0.45	0.09
120910A	0.73	0.04	-0.48	0.05
449342	0.31	0.02	0.19	0.08
449385	0.36	0.02	0.18	0.07
449387	0.73	0.03	-0.09	0.04
452001	0.45	0.09	-0.49	0.14
452013	0.79	0.11	0.12	0.07
449391	0.6	0.03	0.13	0.04
452025	1.49	0.2	-1.1	0.09
461155	0.57	0.03	-0.15	0.05
461107	0.59	0.08	0.78	0.09
530809	0.58	0.12	1.22	0.14
461119	0.58	0.08	0.82	0.09
461143	0.94	0.1	0.25	0.07
451036	0.53	0.1	-0.17	0.1
540443	1.34	0.08	-0.9	0.03
468105	0.63	0.05	-0.55	0.07
452038	1.38	0.14	-0.42	0.04
461131	0.6	0.08	0.85	0.09
530777	0.84	0.06	-0.3	0.05
540418	1.62	0.09	-1.13	0.02
540501	0.94	0.06	-0.76	0.03
540430	1.06	0.11	-1.48	0.09
615920	1.33	0.12	-1.11	0.03
615955	1.36	0.14	-1.63	0.06
615980	2.08	0.15	-1.01	0.02
615940	1.1	0.12	-1.3	0.05
615967	1.68	0.13	-1.03	0.03
659511	1.97	0.22	-1.17	0.04
659703	0.78	0.08	-0.36	0.07
659468	0.91	0.1	0.16	0.08
659535	1.93	0.1	-1.19	0.02
659679	0.55	0.06	-0.02	0.1
659691	1.02	0.05	-0.67	0.03
677531	0.75	0.09	-0.96	0.1
677555	0.75	0.09	-1.15	0.11
677543	1.36	0.14	-0.95	0.06
677567	0.92	0.1	-0.86	0.08
659523	1.66	0.08	-1.11	0.02
659626	0.52	0.06	-0.53	0.11
659643	0.81	0.09	0.47	0.07
659667	0.97	0.13	-0.07	0.08
659715	0.73	0.07	-0.29	0.07
659426	0.69	0.07	-0.15	0.08
659480	0.68	0.08	-0.73	0.09
659492	0.9	0.12	-0.29	0.12
659547	1.46	0.18	-1.35	0.07
659559	1.86	0.09	-1.01	0.02
659602	0.61	0.07	-0.52	0.09
659614	0.93	0.09	-0.33	0.06
677448	0.64	0.1	0.45	0.11
677500	1.39	0.16	-1.26	0.07
677579	1.29	0.13	-0.81	0.06
677515	0.72	0.09	-1.17	0.11
791200	0.54	0.1	1.25	0.13
791161	1	0.08	-0.26	0.06
791184	1.08	0.13	-1.56	0.1
791146	1.22	0.11	-1.02	0.05

Table I-4. IRT Parameters for ELA Grade 6

IREF	a	SE (a)	b	SE (b)
120012A	0.51	0.05	0.2	0.06
120011A	0.32	0.04	1.13	0.14
121328A	1.38	0.1	-1.05	0.04
121329A	3.11	0.26	-0.68	0.02
121359A	1.08	0.07	0	0.03
114439A	1.52	0.11	-1.02	0.04
120013A	0.42	0.09	1.1	0.22
120014A	0.33	0.04	0.15	0.09
121327A	2.5	0.23	-0.76	0.02
121358A	0.93	0.06	0.03	0.04
452219	0.83	0.11	-0.11	0.12
452269	1.99	0.09	-0.38	0.02
452311	0.85	0.06	0.09	0.04
452348	1.16	0.07	-0.1	0.03
452299	1.12	0.08	-0.37	0.04
452231	0.76	0.04	0.11	0.04
452243	0.47	0.04	0.14	0.06
530858	0.64	0.07	0.55	0.07
452257	0.83	0.05	0.22	0.04
452335	0.68	0.09	1.59	0.11
452360	1.38	0.08	-0.25	0.03
530821	1.26	0.07	-0.68	0.03
540986	0.51	0.06	1.19	0.09
541060	0.65	0.07	0.92	0.06
541043	0.95	0.09	0.29	0.06
541091	0.81	0.07	-0.16	0.06
541015	0.73	0.08	0.14	0.08
608280	1.1	0.08	-0.31	0.04
616110	2.12	0.2	-0.67	0.03
616074	1.72	0.16	-0.9	0.05
620737	0.77	0.13	1	0.09
608268	1.16	0.07	-0.68	0.04

IREF	a	SE (a)	b	SE (b)
616086	1.55	0.17	-0.74	0.03
616098	1.42	0.13	-0.93	0.06
616136	1.84	0.19	-0.75	0.03
616272	0.47	0.05	0.49	0.1
616309	0.55	0.06	-0.12	0.08
616347	0.4	0.05	-0.2	0.11
661656	0.62	0.07	-0.47	0.1
661814	2.24	0.21	-0.7	0.02
661620	1.06	0.1	-0.38	0.06
661632	1.49	0.14	-0.72	0.05
661644	1.02	0.11	-0.78	0.07
661668	1.51	0.14	-0.66	0.05
661838	1.21	0.11	-0.55	0.05
661790	3.08	0.26	-0.76	0.02
661802	0.97	0.1	-0.95	0.08
661826	1.76	0.19	-0.94	0.05
677939	0.97	0.12	0.21	0.07
775862	1.22	0.1	-0.32	0.04
782042	1.01	0.09	-0.34	0.05
775847	0.57	0.09	1.01	0.14
775824	0.85	0.11	0.74	0.09
775806	0.55	0.07	0.27	0.09
791241	0.85	0.07	0.35	0.06
775831	0.65	0.07	-0.24	0.08
782057	0.82	0.09	-1.07	0.11
782072	0.78	0.08	-0.82	0.09
782064	0.92	0.08	-0.4	0.05
791220	1.44	0.12	-0.48	0.04
782035	0.73	0.1	-0.21	0.09
782050	0.76	0.09	-0.99	0.1
791248	0.89	0.06	0.61	0.05
791213	1.59	0.12	-0.33	0.03

Table I-5. IRT Parameters for ELA Grade 7

IREF	a	SE (a)	b	SE (b)
115432A	0.54	0.06	-0.18	0.09
114594A	0.68	0.04	-0.67	0.05
121313A	0.74	0.07	-0.46	0.06
121315A	0.68	0.12	0.03	0.09
121997A	0.73	0.04	-0.39	0.04
114593A	1.3	0.06	-0.86	0.03
114596A	0.8	0.04	-0.61	0.04
127690A	0.88	0.09	0.44	0.06
127693A	1.48	0.15	-0.14	0.07
127695A	1.34	0.16	-0.41	0.1
115433A	0.99	0.07	-0.65	0.03
127691A	0.64	0.04	-1.09	0.07
127694A	0.82	0.11	-0.55	0.16
127692A	0.95	0.11	-0.29	0.11
452468	0.67	0.03	-0.13	0.04
452456	0.89	0.13	-0.15	0.06
452480	1.34	0.16	-0.52	0.06
452492	0.79	0.13	-0.76	0.12
537337	0.65	0.05	-0.15	0.07
537369	0.86	0.04	-0.75	0.04
531738	0.57	0.1	1.01	0.11
531750	0.6	0.05	-0.02	0.06
531700	1.04	0.05	-0.98	0.03
537321	1.38	0.11	-0.77	0.03
537297	1.4	0.06	-0.76	0.02
537309	1.51	0.07	-0.72	0.02
610008	0.65	0.08	0.1	0.09
616633	0.84	0.09	0.3	0.07
616718	0.97	0.1	-0.09	0.09
616749	0.89	0.09	0.05	0.08
616773	0.62	0.08	0.09	0.11
616785	0.84	0.09	-0.08	0.1
616799	0.95	0.1	-0.15	0.09
609676	1.16	0.1	-1.05	0.06
616691	0.39	0.06	1.93	0.2
662021	0.7	0.07	-0.61	0.08
662057	0.78	0.11	0.08	0.1
662142	1.47	0.15	-1.02	0.06
616706	0.49	0.07	0.86	0.09
616811	0.44	0.06	0.58	0.1
618684	1.07	0.11	0.09	0.07
662154	1.23	0.11	-0.76	0.06
662081	0.79	0.07	-0.24	0.07
662106	1.46	0.11	-0.99	0.03
662118	0.74	0.08	-1.01	0.1
662045	1.07	0.15	-0.25	0.06
662069	0.97	0.15	-0.63	0.09
662130	1.78	0.18	-0.93	0.05
662033	0.58	0.11	1.2	0.14
678083	0.91	0.09	-0.76	0.07
678095	0.78	0.11	-0.1	0.11
678119	0.63	0.03	-0.35	0.04
775875	0.94	0.11	-1.37	0.1
776002	0.4	0.05	-0.25	0.11
775981	0.65	0.06	-0.09	0.08
775988	0.76	0.16	0.21	0.15
775882	1.14	0.14	-1.51	0.1
775889	1.86	0.16	-0.78	0.03
775910	1.23	0.13	-1.26	0.07
775974	0.83	0.07	-0.31	0.06
776016	0.56	0.06	0.7	0.12
776009	0.61	0.06	0.2	0.09
793031	0.8	0.05	-0.21	0.04
793091	1.24	0.1	-0.67	0.05

Table I-6. IRT Parameters for ELA Grade 8

IREF	a	SE (a)	b	SE (b)
114876A	0.84	0.13	-0.6	0.08
114877A	0.43	0.03	0.4	0.07
114879A	1.02	0.05	-0.55	0.03
122082A	0.72	0.12	-0.54	0.1
449900	0.98	0.04	-0.43	0.03
531810	2.07	0.22	-1.28	0.05
537726	0.75	0.05	0.32	0.04
537740	0.87	0.06	0.08	0.04
538821	0.47	0.1	0.56	0.18
538833	0.98	0.12	-0.37	0.05
538845	1.11	0.13	-0.39	0.04
538857	1.09	0.13	-0.35	0.04
538809	1.13	0.13	-0.54	0.05
537758	0.55	0.05	-0.39	0.08
537774	1	0.06	-0.13	0.04
612180	0.65	0.07	-1.17	0.13
612192	1.77	0.08	-0.74	0.02
612205	0.52	0.06	-1.07	0.14
612217	1.23	0.09	-0.76	0.05
610020	1.98	0.19	-1.15	0.04
612229	1.4	0.07	-0.79	0.02
617052	1.57	0.14	-0.76	0.03
617075	0.58	0.08	0.13	0.09
617908	0.9	0.06	0.21	0.04
617932	1	0.12	-0.2	0.09
617944	0.76	0.05	0.42	0.05
617023	1.94	0.15	-0.91	0.02
617114	0.73	0.08	0.59	0.07
617956	1.06	0.11	0.06	0.07
617144	0.63	0.06	-0.21	0.07
612145	0.7	0.06	-0.87	0.09
617007	1.91	0.2	-1.27	0.05
617038	1.28	0.12	-1.23	0.07
617102	0.72	0.06	-0.32	0.06
617920	0.79	0.06	-0.14	0.05
662270	0.79	0.1	-0.08	0.08
662294	0.6	0.12	0.76	0.14
662343	1.47	0.15	-0.9	0.06
662367	1.37	0.17	-1.15	0.08
662452	0.62	0.03	-0.24	0.04
662258	0.88	0.1	0.21	0.07
662355	1.53	0.21	-1.3	0.1
662476	0.46	0.08	-0.15	0.14
662282	0.49	0.08	0.25	0.12
662306	0.43	0.07	0.35	0.14
662379	1.45	0.14	-0.73	0.05
662391	1.05	0.11	-0.92	0.08
662440	0.51	0.07	-0.04	0.1
678251	1.23	0.05	-0.4	0.02
678263	0.96	0.12	-0.27	0.08
678239	1.34	0.06	-0.86	0.03
662416	0.62	0.03	-0.09	0.04
662428	1.01	0.05	-0.21	0.03
662464	0.48	0.03	0.21	0.06
776534	0.93	0.07	-0.37	0.05
776697	1.54	0.14	-0.83	0.04
776516	0.87	0.07	-0.05	0.05
776525	0.6	0.06	-0.11	0.07
776676	2.99	0.42	-0.47	0.04
776683	1.78	0.18	-0.94	0.04
803879	0.77	0.07	-0.37	0.06
776541	0.95	0.08	-0.6	0.05
776690	1.41	0.15	-0.99	0.05
776669	2.16	0.23	-0.98	0.04

Table I-7. IRT Parameters for ELA Grade HS

IREF	a	SE (a)	b	SE (b)
113699A	0.59	0.04	0.3	0.06
113695A	0.81	0.05	-0.28	0.04
114166A	0.81	0.08	1.09	0.07
121742A	1.32	0.08	-0.55	0.03
121702A	0.74	0.1	-0.23	0.06
121703A	1.22	0.13	-0.58	0.04
121711A	0.88	0.09	0.21	0.06
121714A	1.03	0.1	0.12	0.06
121875A	0.61	0.1	-0.62	0.08
114193A	0.79	0.11	-0.46	0.06
121741A	0.89	0.05	-0.51	0.04
121695A	0.94	0.09	0.6	0.05
453060	0.85	0.08	-0.53	0.04
453074	1.16	0.09	-0.56	0.03
453111	0.69	0.04	0.26	0.04
453087	1.23	0.09	-0.49	0.03
453099	0.77	0.04	-0.08	0.04
124328A	1.02	0.07	0.16	0.04
538901	1.78	0.15	-1.01	0.04
538914	1.4	0.12	-1.04	0.05
538950	0.96	0.06	0.08	0.04
539001	0.61	0.08	1.24	0.1
453138	0.74	0.05	-0.69	0.05
531943	0.6	0.03	0.02	0.04
538889	1.64	0.18	-0.96	0.03
538962	0.74	0.05	0.01	0.04
538986	0.92	0.06	0.12	0.04
539058	0.88	0.05	-0.17	0.03
538926	1.65	0.19	-0.76	0.04
539072	0.99	0.05	-0.66	0.03
538974	0.83	0.06	-0.35	0.05
539013	0.98	0.06	-0.52	0.04
539025	0.92	0.05	-0.42	0.03
618287	1.05	0.11	-0.04	0.07
618316	0.64	0.08	-0.08	0.11
618234	0.86	0.09	1.3	0.08
618275	0.96	0.13	0.74	0.07
612940	0.89	0.07	-0.32	0.05
618247	0.69	0.08	0.58	0.06
662604	0.51	0.04	-0.19	0.05
662616	0.5	0.04	-0.17	0.05
678493	1.19	0.13	-0.72	0.06
678553	1.16	0.14	-1.05	0.09
678420	0.59	0.08	0.29	0.1
678505	1.4	0.15	-0.83	0.06
662568	0.48	0.04	-1.22	0.09
678395	0.83	0.05	-0.87	0.05
678432	1.05	0.1	-0.28	0.06
662580	0.48	0.04	-0.45	0.06
662628	0.67	0.04	0.02	0.04
678469	1	0.05	-0.34	0.03
678517	1.05	0.13	-1	0.09
678529	0.48	0.07	-0.34	0.11
678541	1.35	0.17	-1.06	0.08
777897	1.05	0.09	-0.39	0.05
777911	0.68	0.07	-0.34	0.07
777932	1.08	0.1	-0.65	0.05
778037	1.7	0.18	-0.98	0.05
777918	0.58	0.06	0.27	0.09
778000	2.3	0.26	-1.02	0.04
778023	1.61	0.15	-0.87	0.04
778030	0.55	0.07	-1.2	0.15
777904	1.01	0.11	-0.04	0.06
777951	2.27	0.25	-0.99	0.04
778051	1.52	0.17	-1.07	0.06
777958	2.31	0.26	-1.01	0.04
777967	2.04	0.22	-0.99	0.04
777983	2.55	0.27	-0.95	0.03
778044	1.09	0.14	-1.36	0.11
793164	0.88	0.06	-0.13	0.04
793157	1.11	0.06	-0.36	0.03
793199	1.67	0.17	-0.95	0.05
793185	1.55	0.14	-0.83	0.04

Table I-8. IRT Parameters for Mathematics Grade 3

IREF	a	SE (a)	b	SE (b)
110876A	0.68	0.08	0.71	0.07
110842A	0.75	0.04	-0.48	0.04
110855A	0.73	0.03	0.13	0.03
110975A	0.58	0.05	0.21	0.06
110923A	1.34	0.08	-0.74	0.03
111010A	0.76	0.08	-0.9	0.09
111387A	0.76	0.09	0.64	0.07
111377A	0.91	0.07	-0.45	0.04
111386A	0.63	0.04	0.11	0.05
111420A	1	0.1	0.47	0.06
111426A	0.7	0.05	0.4	0.05
111405A	1.3	0.11	-0.95	0.04
111425A	0.61	0.08	-0.55	0.09
111390A	0.62	0.06	-0.39	0.06
111391A	0.61	0.08	0.3	0.09
111429A	0.61	0.09	-0.48	0.08
111649A	0.64	0.08	-1.3	0.14
112559A	0.8	0.09	-0.67	0.06
111650A	1.01	0.07	-0.22	0.04
112569A	1.06	0.1	-0.83	0.05
112579A	0.67	0.06	-0.98	0.09
112552A	0.77	0.1	-0.32	0.14
112553A	0.68	0.12	0.34	0.12
112595A	0.7	0.04	-0.28	0.04
112601A	1.07	0.11	0.05	0.07
112600A	1.23	0.11	-0.6	0.05
112615A	0.6	0.07	0.88	0.09
120682A	0.83	0.1	-0.28	0.11
442035	0.7	0.07	0.42	0.07
442166	0.71	0.09	0.86	0.09
442037	0.95	0.05	-0.31	0.03
442416	1.26	0.12	-0.59	0.04
451098	0.82	0.06	-1.03	0.05
451116	0.74	0.08	1	0.07
463081	1.23	0.06	0.18	0.03
528813	0.6	0.04	-0.35	0.05
529126	0.58	0.04	-0.33	0.05
529146	0.59	0.09	0.03	0.1
597535	1	0.1	0.32	0.06
597576	0.79	0.09	0.41	0.07
598262	0.79	0.09	0.57	0.08
598375	0.59	0.11	1.45	0.16
603958	1	0.12	0.82	0.13
604382	0.95	0.1	0.71	0.06
598429	0.8	0.09	-0.99	0.06
604123	1.11	0.06	0.74	0.04
598100	0.75	0.14	1	0.11
656824	0.61	0.08	-0.68	0.11
656836	0.52	0.04	-0.77	0.06
656868	1.5	0.14	0.04	0.04
656874	0.75	0.09	-0.23	0.08
595744	1.16	0.11	0.52	0.05
597702	0.5	0.09	-0.35	0.15
597774	1.39	0.17	-0.69	0.06
598213	0.6	0.1	-0.71	0.18
656872	0.52	0.08	-0.98	0.15
656826	0.6	0.07	-0.09	0.1
656884	1.49	0.12	-0.97	0.04
656860	0.81	0.1	-1.04	0.11
656876	0.88	0.12	0.26	0.08
656892	1.47	0.13	0	0.05
773480	0.7	0.07	1.14	0.13
773475	0.73	0.14	1.7	0.2
773556	0.42	0.08	0.91	0.19
773562	0.73	0.08	-0.83	0.08
773530	1.01	0.1	0.99	0.1
773532	0.66	0.07	-0.15	0.07
773536	0.51	0.06	-1.15	0.13
773546	0.66	0.07	-0.97	0.1

Table I-9. IRT Parameters for Mathematics Grade 4

IREF	a	SE (a)	b	SE (b)
111136A	1.2	0.08	-0.22	0.03
111166A	0.95	0.05	0.18	0.03
111122A	0.74	0.12	-0.58	0.08
111123A	1.17	0.08	-0.19	0.03
111162A	1.29	0.16	0.01	0.09
111663A	1.05	0.06	-0.67	0.04
111681A	1.89	0.17	0.61	0.03
111712A	0.61	0.04	0.43	0.05
111730A	0.75	0.09	0.25	0.08
111731A	0.79	0.13	0.28	0.09
111686A	0.68	0.04	0.27	0.04
111687A	1.36	0.14	0.58	0.04
111707A	0.68	0.1	1.24	0.1
111716A	0.83	0.09	0.06	0.06
111717A	0.73	0.04	-0.16	0.04
111685A	0.98	0.11	-0.71	0.05
111696A	1.4	0.12	0.11	0.03
111658A	1	0.11	-0.66	0.07
111678A	1.43	0.12	0.37	0.03
111722A	0.86	0.1	0.9	0.07
112838A	0.78	0.1	0.55	0.07
112817A	0.94	0.11	1.12	0.07
112837A	0.65	0.1	1.12	0.1
112794A	0.92	0.07	0.05	0.04
112783A	1.28	0.12	0.2	0.04
112812A	0.79	0.1	1.3	0.09
120551A	0.85	0.11	1.34	0.09
122153A	1.11	0.13	0.19	0.06
121750A	1.36	0.09	-0.47	0.04
446780	0.54	0.1	0.38	0.22
445558	1.23	0.13	0.78	0.04
445572	0.83	0.05	-0.51	0.04
445588	0.81	0.08	0	0.06
446443	0.6	0.04	-0.63	0.06
463067	0.99	0.1	-0.24	0.04
463027	1.06	0.12	0.96	0.06
454751	0.82	0.05	0.25	0.04
455024	1.01	0.05	-0.55	0.03
529949	0.56	0.06	0.21	0.08
599462	0.94	0.13	-0.04	0.1
599574	0.46	0.06	0.98	0.12
599952	0.93	0.12	-0.2	0.09
599432	0.74	0.08	0.17	0.06
599954	0.66	0.09	0.98	0.12
600599	0.63	0.09	1.17	0.14
599071	0.98	0.1	-0.86	0.08
600647	0.66	0.12	0.15	0.18
600654	0.55	0.06	-0.18	0.07
656926	0.76	0.05	-0.35	0.04
600620	0.87	0.06	-0.63	0.04
600671	0.65	0.1	0.58	0.08
656932	0.83	0.11	-0.69	0.06
656960	1.11	0.14	-0.93	0.1
656896	0.61	0.09	-0.36	0.1
656924	0.6	0.08	0.28	0.1
656946	1.58	0.16	0.11	0.04
656954	1.28	0.14	0.22	0.05
656920	0.64	0.1	-0.61	0.11
774522	0.7	0.08	0.44	0.08
774534	0.65	0.07	-0.38	0.08
774548	0.95	0.1	-0.72	0.07
774560	0.64	0.17	1.16	0.18
774506	0.96	0.09	-0.11	0.05
774536	0.55	0.07	-0.42	0.09
774540	0.52	0.07	0.37	0.1
774542	0.91	0.1	0.81	0.09
774562	0.91	0.09	-0.59	0.07

Table I-10. IRT Parameters for Mathematics Grade 5

IREF	a	SE (a)	b	SE (b)	IREF	a	SE (a)	b	SE (b)
111242A	1.36	0.13	-0.73	0.04	450200	1.41	0.14	-0.48	0.05
111257A	0.69	0.08	-1.21	0.13	450352	0.55	0.08	-0.58	0.1
111294A	1.36	0.12	0.35	0.04	449994	0.66	0.06	0.09	0.06
111275A	0.57	0.09	-0.69	0.08	450339	0.89	0.09	0.56	0.06
111243A	0.87	0.04	0.23	0.04	449975	0.9	0.07	0.17	0.05
111277A	0.72	0.06	0.25	0.05	450210	1.25	0.12	-0.74	0.04
112346A	0.8	0.04	-0.61	0.04	450296	0.56	0.06	1.24	0.11
112354A	0.56	0.07	1.21	0.1	532983	0.98	0.07	0.83	0.06
112364A	0.47	0.05	1.18	0.11	532547	0.76	0.1	0.78	0.13
112373A	0.68	0.08	0.14	0.09	532745	0.57	0.1	0.51	0.13
112358A	0.88	0.06	0.55	0.04	601957	0.55	0.06	-0.09	0.08
112384A	1.07	0.06	0.35	0.03	601986	1.36	0.13	-0.96	0.04
112408A	1.01	0.11	0.44	0.07	604399	0.52	0.06	0.33	0.09
112410A	1.05	0.12	0.35	0.06	601619	1.11	0.12	-0.7	0.05
112335A	1.18	0.12	-0.72	0.05	602047	0.71	0.06	0.3	0.07
112348A	0.76	0.08	0.63	0.06	602060	0.68	0.1	-0.97	0.08
112353A	0.64	0.06	0.26	0.08	602432	0.63	0.1	-1.1	0.09
112368A	1.09	0.11	0.29	0.05	657008	0.73	0.12	0.16	0.09
112372A	1.09	0.05	-0.77	0.03	657018	0.72	0.11	0.92	0.13
112377A	0.62	0.04	-0.77	0.05	657036	0.55	0.09	0.73	0.14
112385A	0.78	0.05	0.39	0.04	604407	0.75	0.05	1.48	0.09
113843A	1.31	0.14	-0.37	0.06	657002	0.64	0.1	1.12	0.16
113859A	1.02	0.08	-0.84	0.06	657038	0.95	0.14	1.48	0.17
113884A	0.55	0.04	-0.73	0.06	656974	0.97	0.12	-1.1	0.1
112392A	0.94	0.09	1.22	0.07	657022	0.95	0.11	-0.7	0.06
113872A	0.94	0.09	0.86	0.06	656968	0.66	0.09	-1.08	0.14
113877A	0.53	0.08	1.48	0.15	657010	0.6	0.08	0.13	0.08
113889A	1.64	0.15	-0.75	0.03	774366	0.58	0.14	0.11	0.19
120737A	0.96	0.05	0.74	0.03	774374	0.57	0.07	0.64	0.12
121515A	1.36	0.25	0.56	0.19	774404	0.86	0.09	1.13	0.12
113862A	0.82	0.1	0.64	0.1	774436	1.13	0.11	-0.85	0.06
120730A	0.88	0.12	1.1	0.16	774441	0.58	0.09	0.61	0.11
450153	1	0.05	-0.77	0.04	774378	0.57	0.06	0.51	0.1
450180	0.64	0.07	1.47	0.11	774414	0.52	0.07	-1.03	0.13
450194	1.05	0.1	0.47	0.05	774422	0.77	0.08	0.07	0.07

Table I-11. IRT Parameters for Mathematics Grade 6

IREF	a	SE (a)	b	SE (b)
110905A	1.03	0.16	-0.81	0.06
110982A	0.75	0.08	0.9	0.06
110986A	0.73	0.04	-0.19	0.04
110941A	0.93	0.08	-0.44	0.06
111025A	1.04	0.08	-0.2	0.03
111035A	0.74	0.11	0.08	0.07
111445A	1.25	0.11	0.35	0.05
111465A	0.69	0.08	0.43	0.08
111488A	1.32	0.16	-0.33	0.06
111514A	0.79	0.04	-0.19	0.04
111518A	0.8	0.05	0.33	0.04
111455A	1.1	0.11	0.18	0.06
111479A	1.26	0.12	0.08	0.04
111630A	1.06	0.07	-1.05	0.06
112653A	1.1	0.12	-0.22	0.04
112656A	1.36	0.19	-0.78	0.05
112655A	0.78	0.08	0.85	0.06
112699A	0.97	0.09	0.58	0.05
112672A	1.08	0.06	-0.03	0.03
112673A	0.92	0.07	-0.25	0.04
112666A	0.94	0.08	-0.69	0.07
112697A	1.82	0.15	-0.3	0.03
120494A	0.8	0.04	-0.18	0.04
442369	0.73	0.07	-0.03	0.05
442631	0.78	0.07	0.21	0.07
442813	1.19	0.13	-0.78	0.07
442538	1.05	0.09	0.75	0.05
120854A	0.83	0.05	0.17	0.04
442634	0.82	0.08	-0.95	0.09
442683	1.02	0.14	-0.33	0.07
450365	0.73	0.07	-0.01	0.06
450436	1.29	0.18	-0.83	0.05
453755	1.24	0.11	0.5	0.04
453771	0.9	0.04	0.1	0.03
453780	1.14	0.08	-0.84	0.04
534783	0.99	0.11	-0.02	0.05
442356	1.17	0.11	0.3	0.05
442566	1.31	0.06	-0.56	0.03
442628	0.67	0.04	0.48	0.05
442641	1.01	0.1	0.18	0.07
450368	0.68	0.06	-0.24	0.04
453675	0.96	0.12	-0.74	0.05
534823	1.21	0.15	-0.17	0.09
603290	1.21	0.08	-0.19	0.03
603418	1.45	0.14	0.01	0.06
603458	1.04	0.16	-0.9	0.06
602872	1.28	0.1	-0.56	0.04
603283	1.01	0.08	-0.41	0.05
603329	0.68	0.05	0.15	0.05
603349	1.31	0.18	-0.89	0.05
603712	1.37	0.14	-0.8	0.04
657074	1.13	0.11	-0.22	0.05
657114	0.98	0.1	-0.22	0.06
657040	0.94	0.12	-0.93	0.1
657042	1.16	0.14	-0.55	0.05
657062	0.52	0.09	0.39	0.11
657092	1.17	0.13	-1.07	0.05
657054	0.95	0.1	-0.35	0.06
657060	1.59	0.17	-0.6	0.05
773619	0.99	0.1	-0.74	0.07
773720	1.02	0.08	0.18	0.05
773655	0.6	0.06	0.37	0.08
773627	0.66	0.12	1.44	0.15
773697	0.92	0.08	0.66	0.07
773661	0.66	0.07	0.63	0.09
773681	0.84	0.08	0.36	0.07
773685	0.62	0.06	0.81	0.1
773716	1.05	0.11	-0.91	0.08

Table I-12. IRT Parameters for Mathematics Grade 7

IREF	a	SE (a)	b	SE (b)
111104A	0.68	0.05	-0.34	0.05
111048A	1.11	0.11	-0.03	0.05
111106A	1.09	0.1	1.11	0.06
111054A	1.42	0.12	-0.82	0.03
111071A	0.9	0.1	0.11	0.08
111074A	0.88	0.09	-1.05	0.05
111055A	0.61	0.04	-0.07	0.05
111075A	1.39	0.14	-0.04	0.06
111130A	0.76	0.09	-1.03	0.07
111085A	0.63	0.08	0.72	0.09
111092A	1.04	0.1	-1.08	0.04
111099A	0.72	0.05	0.04	0.04
111090A	0.93	0.09	-1.12	0.05
111745A	0.58	0.08	0.04	0.15
111765A	0.9	0.06	0.31	0.04
111774A	0.94	0.07	-0.17	0.05
111778A	0.67	0.09	-1.62	0.11
111795A	0.82	0.05	0.4	0.05
111764A	0.94	0.1	0.23	0.07
111766A	1.34	0.12	0.27	0.05
111804A	0.56	0.05	-0.1	0.07
111769A	0.74	0.04	-0.54	0.04
111783A	0.87	0.12	-1.39	0.14
111796A	0.98	0.1	0.43	0.06
112909A	0.81	0.05	-1.13	0.05
113101A	1.34	0.12	0.33	0.05
112523A	0.77	0.09	0.5	0.07
112881A	0.59	0.03	0.22	0.05
112887A	1.25	0.12	0.32	0.05
446491	0.66	0.04	-1.22	0.06
446589	0.91	0.09	0.59	0.06
446722	1.38	0.13	0.13	0.05
446831	0.67	0.1	0.14	0.09
446901	1.09	0.1	-1.27	0.05
454054	0.87	0.1	-0.87	0.08
530309	0.55	0.08	-0.04	0.08
446543	0.74	0.04	0.33	0.04
446673	0.64	0.04	-1.46	0.07
451883	0.84	0.09	-1.37	0.12
454049	0.8	0.09	0.08	0.09
530662	0.52	0.09	1.42	0.15
531427	1.59	0.15	-0.61	0.04
531205	0.84	0.09	0.11	0.08
604898	1.23	0.11	-0.38	0.04
605009	0.93	0.04	-0.3	0.03
606499	0.7	0.06	-0.01	0.06
605348	0.98	0.13	-0.09	0.11
605643	1.2	0.1	-1.31	0.05
605668	0.71	0.06	0.34	0.07
605075	0.55	0.06	-0.4	0.08
605241	0.85	0.1	-0.28	0.08
605361	0.78	0.11	0.42	0.1
606474	0.59	0.05	0.19	0.08
606510	0.89	0.07	-0.57	0.05
605373	0.84	0.04	-0.02	0.04
606208	0.7	0.06	0.25	0.07
657125	0.6	0.09	0.11	0.1
657178	0.65	0.09	0.66	0.11
657158	0.71	0.1	0.48	0.1
657166	0.9	0.16	1.24	0.13
657168	0.77	0.1	-1.3	0.12
657170	0.69	0.1	-0.08	0.1
657138	0.52	0.09	-0.32	0.15
657140	0.61	0.09	0.14	0.09
774676	0.53	0.06	-0.74	0.1
774659	1.08	0.12	-1.2	0.08
774637	1.34	0.14	-1.08	0.06
774715	0.62	0.06	0	0.08
774713	0.57	0.06	-0.31	0.08
774719	1.06	0.09	-0.35	0.05
774735	0.54	0.11	1.52	0.2

Table I-13. IRT Parameters for Mathematics Grade 8

IREF	a	SE (a)	b	SE (b)
111286A	1.11	0.11	-0.01	0.05
111335A	0.89	0.1	-0.23	0.08
111339A	0.66	0.07	0.88	0.07
111565A	0.86	0.05	-0.73	0.04
111588A	1.13	0.11	-0.27	0.07
111594A	0.89	0.08	0.31	0.06
111622A	0.95	0.06	0.16	0.04
111560A	0.84	0.04	0.14	0.04
111352A	0.79	0.04	-0.87	0.04
112499A	1.47	0.12	-0.88	0.03
112511A	1.18	0.12	0	0.04
111583A	0.69	0.07	0.91	0.06
111615A	0.61	0.03	-0.66	0.05
112467A	0.83	0.04	-0.11	0.03
112474A	1.5	0.11	-1.23	0.04
112506A	1.32	0.12	-0.43	0.05
112516A	0.79	0.04	0.11	0.04
112464A	0.78	0.04	-0.73	0.04
112475A	0.97	0.07	-0.56	0.06
112490A	0.99	0.09	0.14	0.06
113922A	1.02	0.07	-0.11	0.04
113952A	1	0.11	0.06	0.05
112510A	0.89	0.08	0.43	0.06
113959A	0.46	0.05	1.05	0.08
113908A	1.12	0.11	-0.17	0.04
113932A	0.75	0.05	0.26	0.04
113937A	0.67	0.04	0.56	0.04
113963A	0.73	0.04	0.18	0.04
113964A	1.07	0.1	0.25	0.05
120571A	0.92	0.11	0.04	0.06
120572A	0.76	0.07	0.66	0.06
113909A	1.46	0.12	-0.79	0.03
113923A	1.05	0.09	0.22	0.06
113943A	1.47	0.13	-0.13	0.03
120560A	0.98	0.05	-0.01	0.03
447054	0.69	0.03	0.02	0.03
446958	0.62	0.07	0.25	0.09
447063	1.11	0.1	-0.54	0.05
447166	1.47	0.11	-1.13	0.03
446979	0.62	0.1	-0.19	0.07
454069	1.05	0.11	-0.14	0.08
454122	0.82	0.08	-0.75	0.06
447047	0.87	0.08	0.59	0.05
532251	0.81	0.08	0.14	0.08
532339	0.84	0.12	0.47	0.18
454205	0.73	0.04	0.03	0.04
532471	1.62	0.15	0.12	0.04
609162	0.67	0.07	0.71	0.07
610515	0.52	0.1	-0.15	0.09
607045	0.74	0.12	-1.25	0.17
607825	0.43	0.05	-0.26	0.09
608123	1.1	0.11	-0.13	0.04
609804	0.53	0.05	0.13	0.08
606892	0.82	0.04	0.45	0.04
607977	0.62	0.06	-0.61	0.08
608070	0.63	0.06	0.06	0.07
608186	0.72	0.08	-1.24	0.12
609908	0.86	0.09	-0.6	0.06
610058	0.98	0.09	0.24	0.06
657232	0.63	0.1	0.41	0.1
657204	0.61	0.1	-0.46	0.14
657254	0.63	0.07	0.33	0.09
657222	0.62	0.08	-1.15	0.14
657256	0.83	0.11	0.51	0.08
774999	0.87	0.08	0.54	0.07
775013	0.52	0.08	0.47	0.11
774991	0.7	0.06	0.54	0.08
774923	0.56	0.09	0.17	0.1
774995	0.63	0.06	0.2	0.08
775061	0.71	0.07	-0.27	0.07

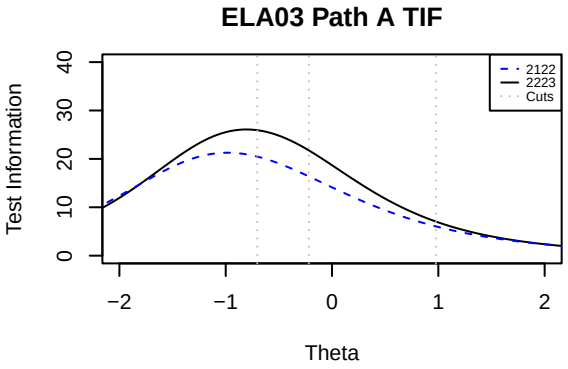
Table I-14. IRT Parameters for Mathematics Grade HS

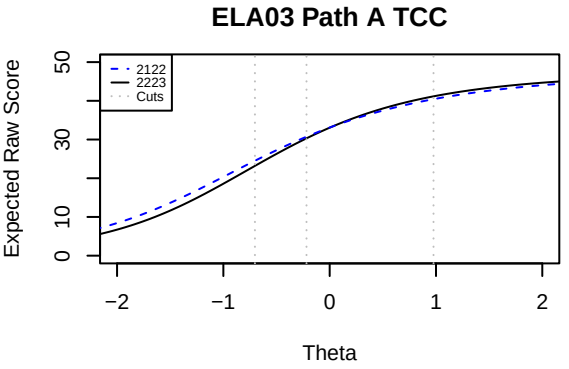
IREF	a	SE (a)	b	SE (b)
110846A	0.91	0.06	0.11	0.04
110867A	0.78	0.1	-1	0.07
110914A	1.43	0.12	0.33	0.04
110881A	0.9	0.05	0.01	0.03
111002A	1.85	0.15	-0.08	0.03
110936A	0.86	0.05	-0.26	0.03
111000A	1.24	0.11	0.23	0.04
110913A	1.01	0.06	0.32	0.03
110921A	1.53	0.13	-0.01	0.03
111033A	0.96	0.17	-0.21	0.1
111016A	1	0.08	0.03	0.04
111534A	1.38	0.12	0.47	0.04
111533A	1.69	0.17	-0.76	0.04
111537A	0.63	0.07	0.96	0.08
111830A	1.14	0.13	-0.88	0.05
111810A	2.14	0.16	0.42	0.03
111829A	1.1	0.1	0.49	0.04
111834A	1.12	0.14	-0.36	0.06
112706A	1.3	0.14	-0.66	0.04
112924A	1.35	0.13	-0.47	0.04
111824A	1.35	0.16	-0.4	0.05
112701A	1.8	0.18	-0.7	0.03
112702A	1.12	0.1	0.39	0.04
112727A	1.1	0.06	-0.13	0.03
112718A	0.91	0.06	0.57	0.04
112744A	0.93	0.09	0.25	0.06
112940A	0.92	0.09	0.49	0.05
112717A	0.79	0.09	0.16	0.08
112722A	1.45	0.1	-0.23	0.03
112934A	0.93	0.11	-0.77	0.08
443381	1.13	0.13	-0.71	0.05
443494	1.7	0.14	0.06	0.03
443515	0.84	0.06	-0.86	0.06
443575	1.28	0.07	-0.58	0.03
454987	1.31	0.09	-0.17	0.03
533908	0.61	0.08	0.85	0.11
533370	0.69	0.08	0.41	0.07
533128	1.05	0.1	0	0.04
613392	0.82	0.09	-0.84	0.09
613828	1.28	0.09	-0.92	0.04
613850	1.06	0.06	-0.67	0.04
613943	0.99	0.14	-0.14	0.09
613955	0.85	0.06	-0.1	0.04
613648	0.86	0.05	-0.2	0.03
614219	1.17	0.1	-0.18	0.04
657330	0.46	0.08	0.49	0.14
613838	2	0.15	0.41	0.03
613961	1.06	0.19	0	0.12
657306	0.78	0.1	-0.37	0.08
657312	1.02	0.13	-0.59	0.08
657320	0.55	0.13	0.94	0.17
657300	0.86	0.13	-0.95	0.12
657266	1.2	0.13	-0.11	0.05
657278	2.28	0.17	0.39	0.03
657284	0.96	0.07	-0.08	0.04
657286	0.48	0.08	0.75	0.16
657290	0.71	0.1	0.02	0.08
657292	0.75	0.1	0.38	0.08
657318	1.14	0.13	-0.66	0.05
775288	0.79	0.1	0.43	0.08
775250	0.61	0.09	0.58	0.11
775312	0.52	0.08	0.55	0.12
775238	0.99	0.1	0.1	0.05
775246	0.79	0.09	0.37	0.07
775254	0.8	0.09	-0.44	0.07
775260	0.68	0.08	0.04	0.07
775316	0.64	0.08	0.02	0.08
775324	0.75	0.15	0.41	0.11

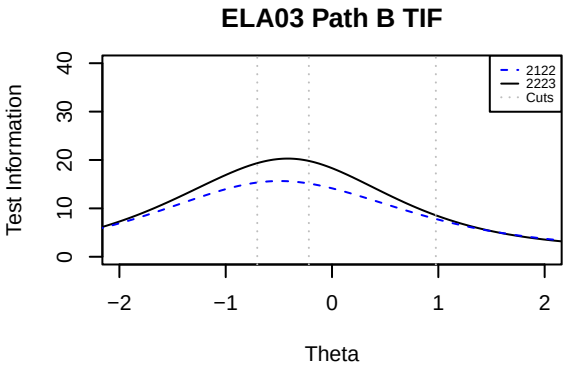
Table I-15. IRT Polytomous Item Parameters for ELA Writing Prompts

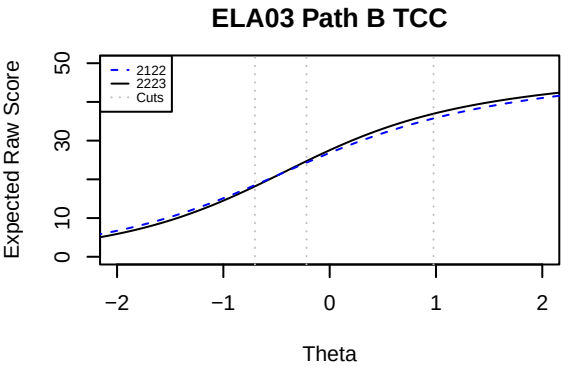
Content Area	Grade	ID	a	b	d0	d1	b-d0	b-d1
ELA–Writing	3	125971C	0.6840	0.5211	1.0771	-1.0771	-0.5561	1.5982
		125971I	0.7126	1.8053	1.0042	-1.0042	0.8011	2.8095
		125971O	0.7834	1.5921	1.6280	-1.6280	-0.0360	3.2201
		464856C	0.7418	-0.0199	0.8126	-0.8126	-0.8325	0.7926
		464856I	0.7059	0.1523	0.6606	-0.6606	-0.5083	0.8129
		464856O	0.6394	0.1897	1.3501	-1.3501	-1.1604	1.5397
	4	126163C	0.7045	0.5089	1.0087	-1.0087	-0.4997	1.5176
		126163I	0.7447	1.7838	0.4723	-0.4723	1.3115	2.2561
		126163O	0.7374	2.0440	1.3395	-1.3395	0.7045	3.3835
		469093C	0.8509	-0.1571	0.7724	-0.7724	-0.9295	0.6153
		469093I	0.7407	0.3089	0.5270	-0.5270	-0.2181	0.8359
		469093O	0.7416	0.3488	1.0937	-1.0937	-0.7448	1.4425
	5	126972C	0.8106	0.3210	0.8910	-0.8910	-0.5700	1.2120
		126972I	0.8932	1.6482	0.8692	-0.8692	0.7790	2.5174
		126972O	0.9110	1.3556	0.7726	-0.7726	0.5830	2.1283
		471924C	0.8713	-0.3914	0.6892	-0.6892	-1.0806	0.2978
		471924I	0.8418	0.3436	0.9387	-0.9387	-0.5951	1.2823
		471924O	0.8010	-0.0543	0.9881	-0.9881	-1.0425	0.9338
	6	127286C	0.8353	0.4033	0.7377	-0.7377	-0.3344	1.1409
		127286I	0.7846	0.8985	0.9704	-0.9704	-0.0719	1.8689
		127286O	0.8372	1.5394	1.3975	-1.3975	0.1419	2.9369
		471934C	0.8835	-0.0342	0.6401	-0.6401	-0.6743	0.6059
		471934I	0.9852	0.0006	0.4699	-0.4699	-0.4693	0.4705
		471934O	0.7926	0.3549	1.2038	-1.2038	-0.8489	1.5587
	7	127658C	0.8392	0.2944	0.8990	-0.8990	-0.6046	1.1934
		127658I	0.8735	1.3833	1.0378	-1.0378	0.3455	2.4211
		127658O	1.0541	1.3370	1.6467	-1.6467	-0.3097	2.9836
		471948C	0.6427	0.1409	0.7419	-0.7419	-0.6010	0.8828
		471948I	0.8558	-0.1643	0.5405	-0.5405	-0.7049	0.3762
		471948O	0.8977	0.2850	1.1199	-1.1199	-0.8349	1.4049
	8	127794C	0.9048	-0.1088	0.8878	-0.8878	-0.9966	0.7790
		127794I	1.2151	0.6522	1.1176	-1.1176	-0.4654	1.7697
		127794O	1.2725	0.6348	1.2348	-1.2348	-0.6000	1.8695
		471958C	0.8053	-0.2635	0.5480	-0.5480	-0.8115	0.2845
		471958I	0.7503	0.4558	0.7508	-0.7508	-0.2950	1.2065
		471958O	0.6934	0.9647	1.3724	-1.3724	-0.4077	2.3371
	HS	126858C	0.8108	0.1786	0.9382	-0.9382	-0.7596	1.1169
		126858I	0.6797	1.6778	1.2315	-1.2315	0.4462	2.9093
		126858O	0.8016	0.9870	1.6437	-1.6437	-0.6567	2.6307
		471963C	0.8138	-0.3419	0.5427	-0.5427	-0.8846	0.2008
		471963I	0.7729	0.5071	0.9274	-0.9274	-0.4203	1.4345
		471963O	0.7428	0.3829	1.1897	-1.1897	-0.8069	1.5726

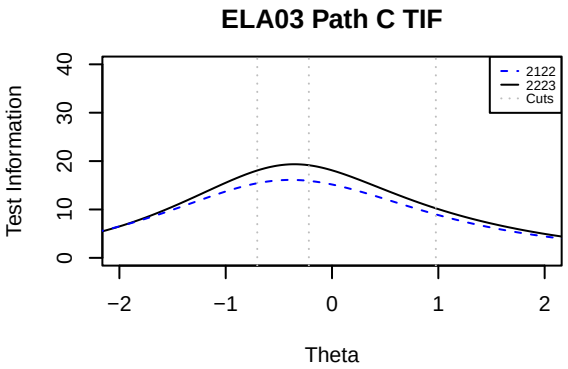
APPENDIX J
TEST CHARACTERISTIC CURVES & TEST INFORMATION FUNCTIONS

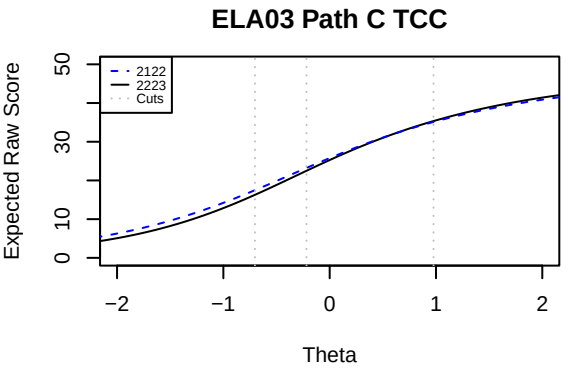

ELA03 Path A TIF
Test Information
Theta
2122
2223
Cuts

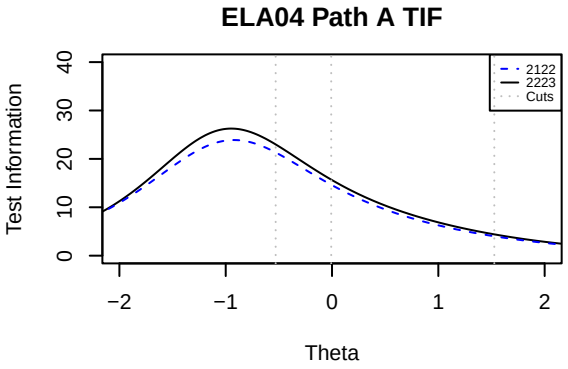

ELA03 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts

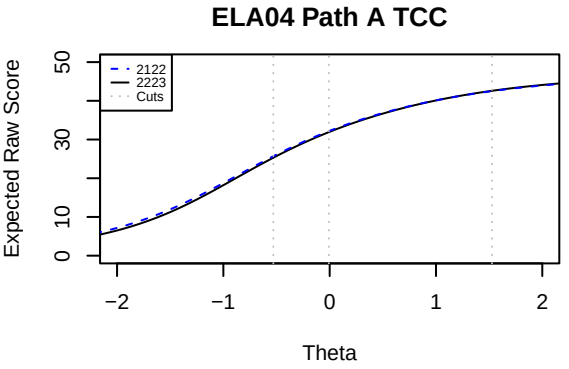

ELA03 Path B TIF
Test Information
Theta
2122
2223
Cuts

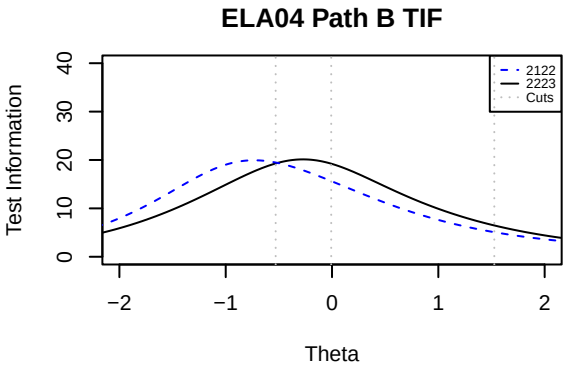

ELA03 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts

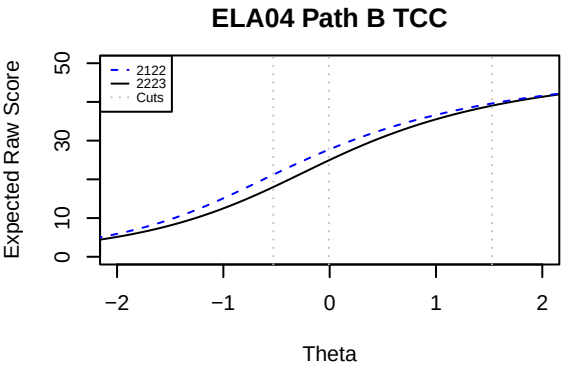

ELA03 Path C TIF
Test Information
Theta
2122
2223
Cuts

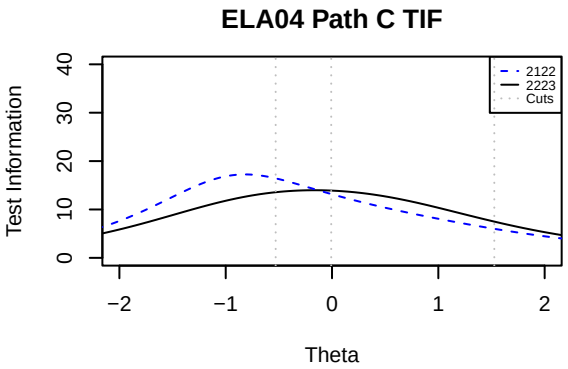

ELA03 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts

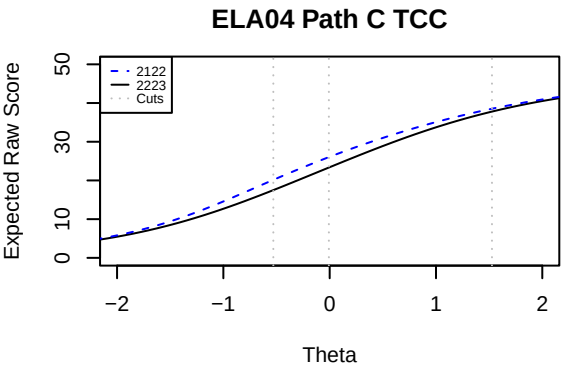

ELA04 Path A TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

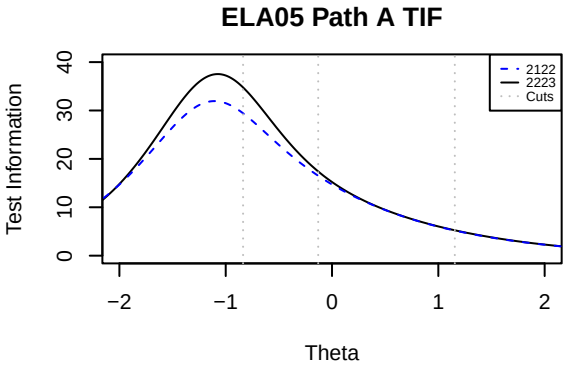

ELA04 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
30
50
−2
−1
0
1
2

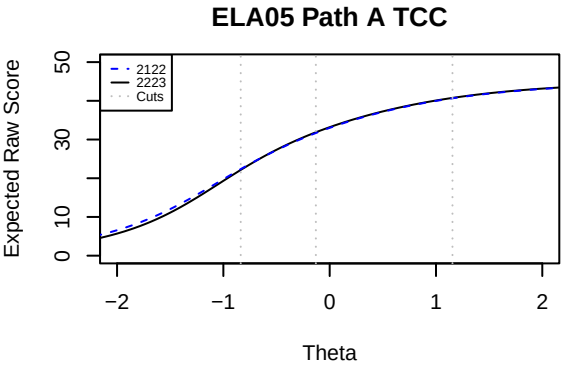

ELA04 Path B TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

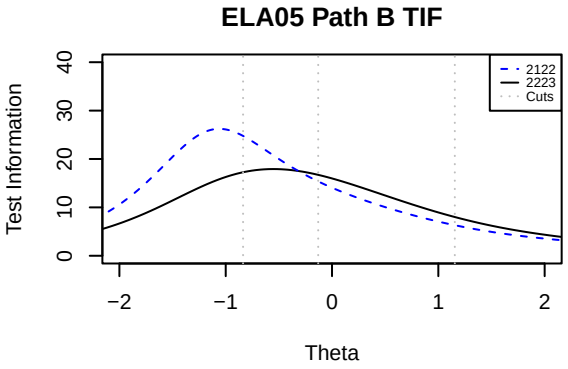

ELA04 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
30
50
−2
−1
0
1
2

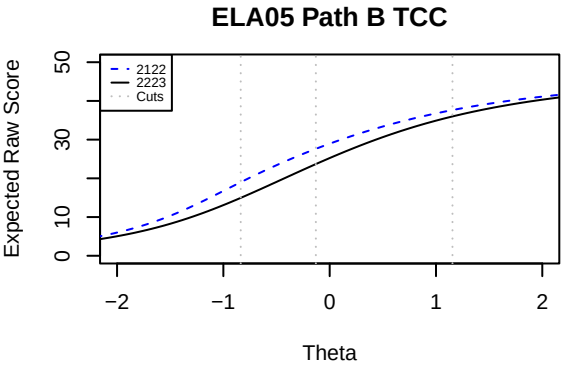

ELA04 Path C TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

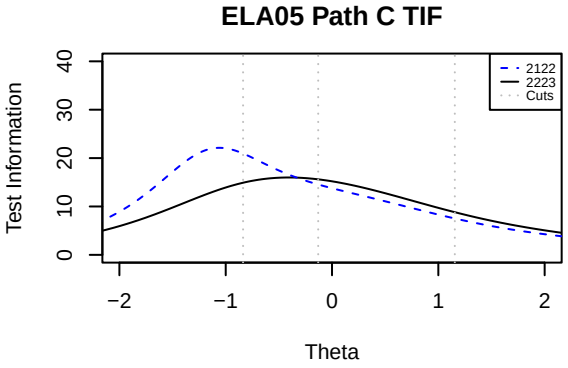

ELA04 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
30
50
−2
−1
0
1
2

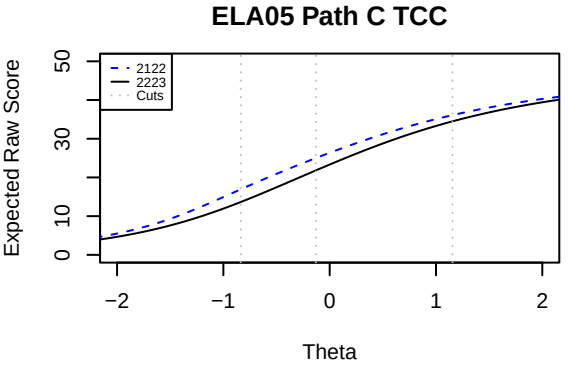

ELA05 Path A TIF
Test Information
Theta
2122
2223
Cuts

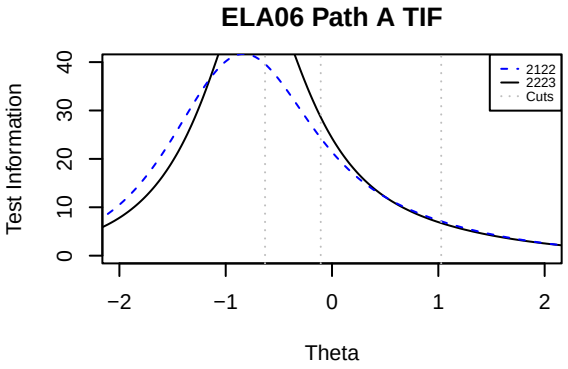

ELA05 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts

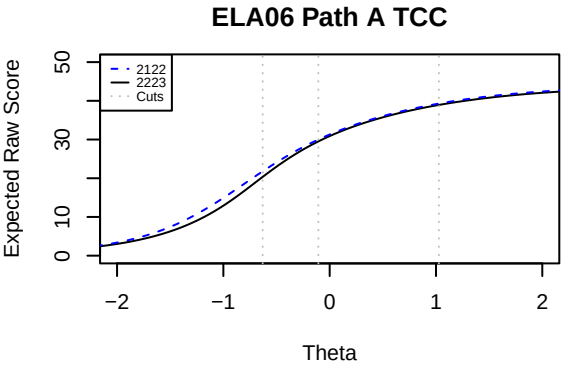

ELA05 Path B TIF
Test Information
Theta
2122
2223
Cuts

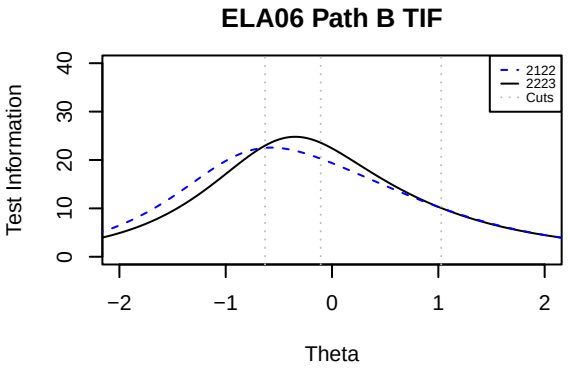

ELA05 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts

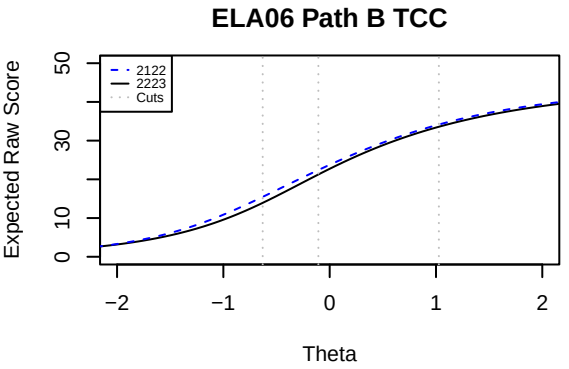

ELA05 Path C TIF
Test Information
Theta
2122
2223
Cuts

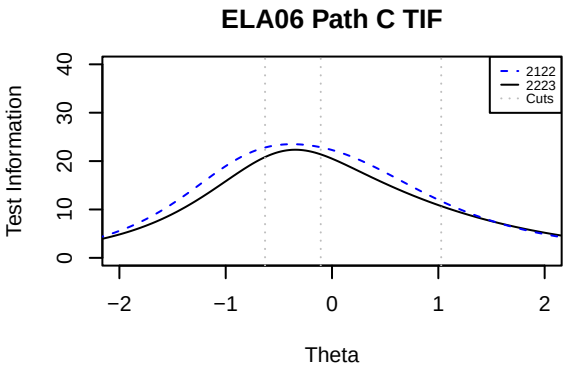

ELA05 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts

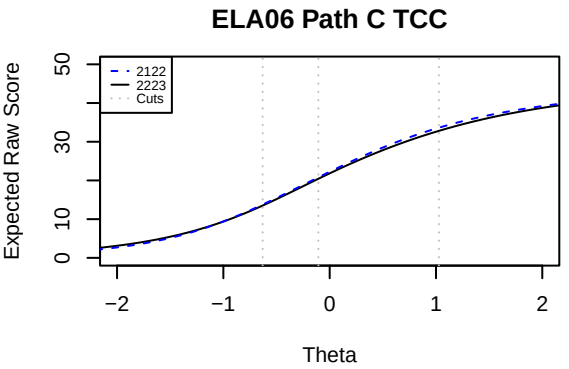

ELA06 Path A TIF
Test Information
Theta
2122
2223
Cuts

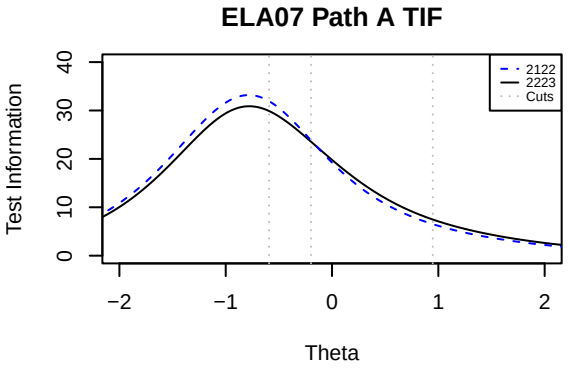

ELA06 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts

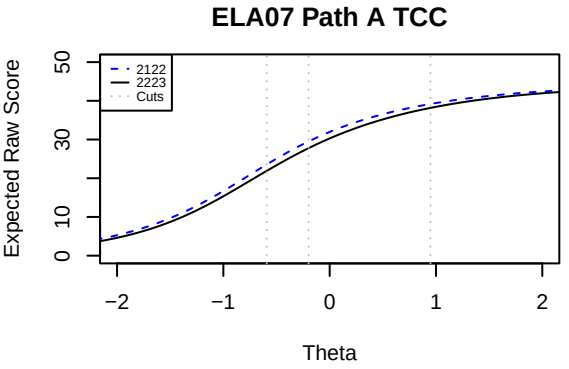

ELA06 Path B TIF
Test Information
Theta
2122
2223
Cuts

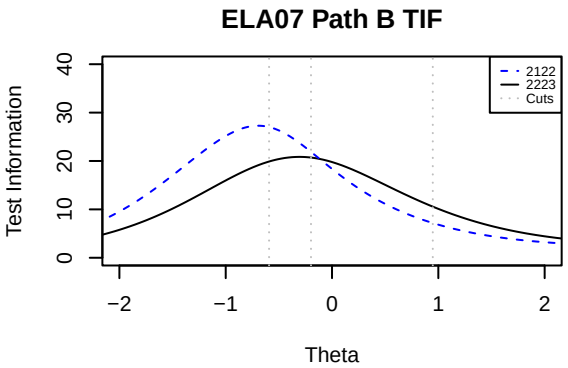

ELA06 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts

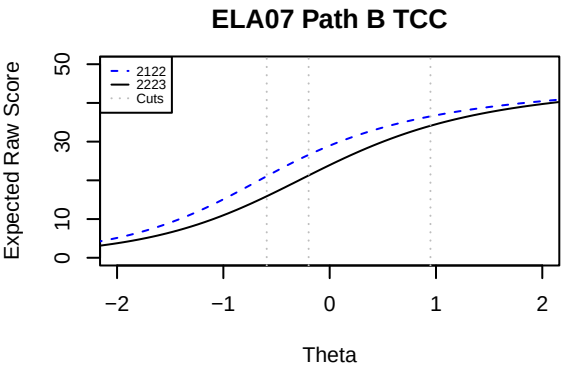

ELA06 Path C TIF
Test Information
Theta
2122
2223
Cuts

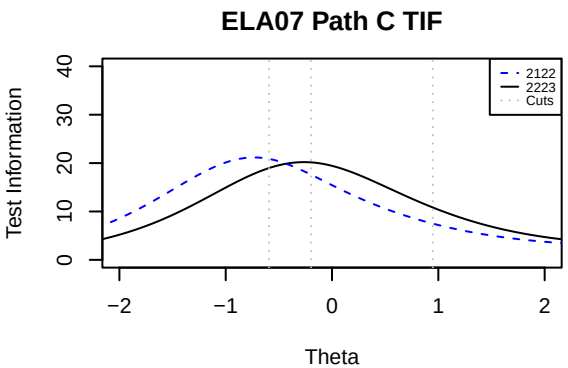

ELA06 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts

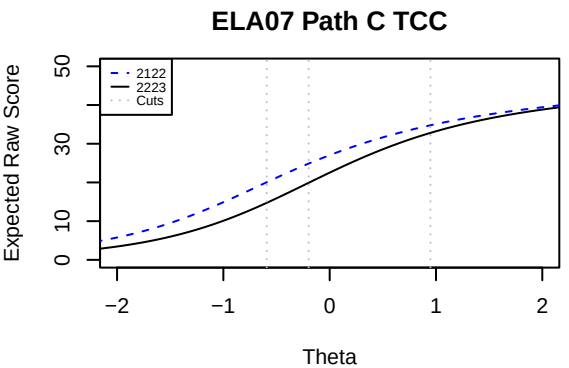

ELA07 Path A TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

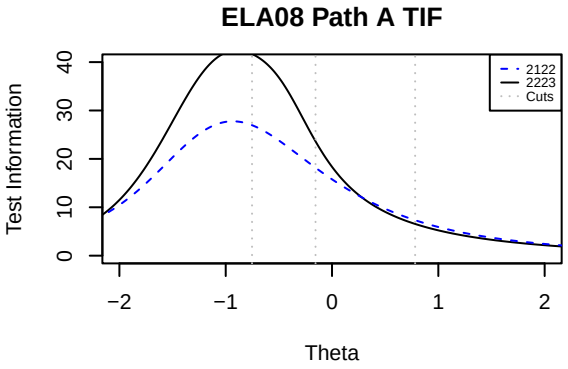

ELA07 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
30
50
−2
−1
0
1
2

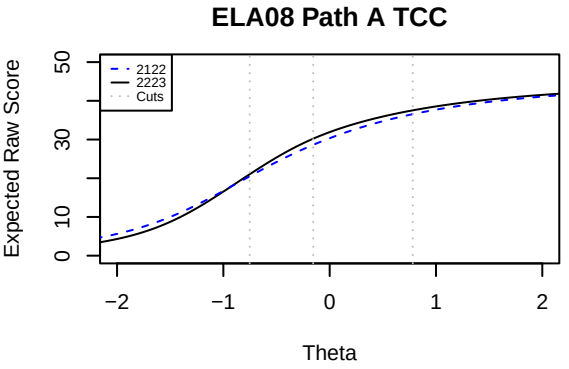

ELA07 Path B TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

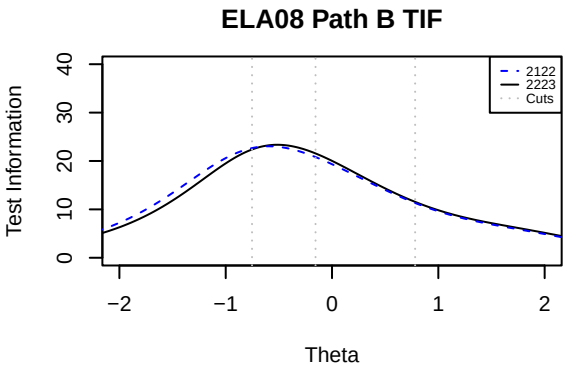

ELA07 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
30
50
−2
−1
0
1
2

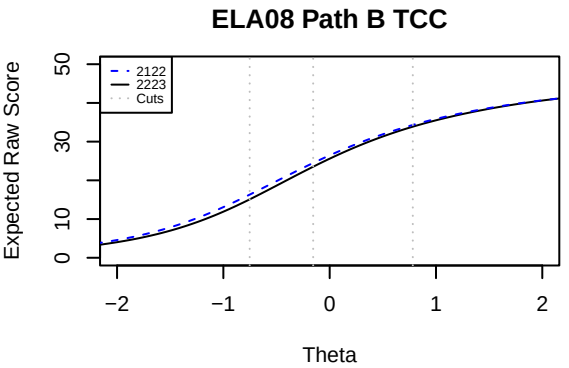

ELA07 Path C TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

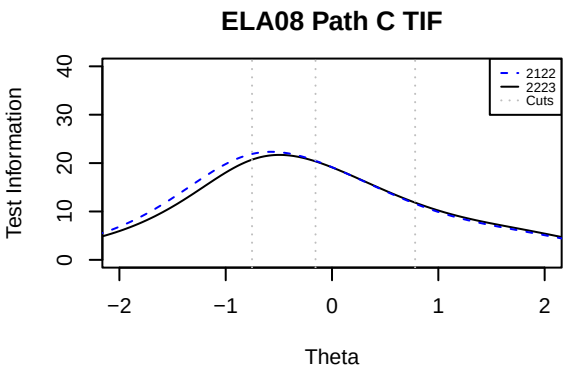

ELA07 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
30
50
−2
−1
0
1
2

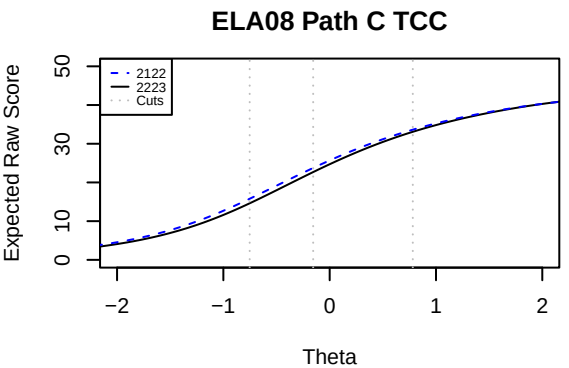

ELA08 Path A TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

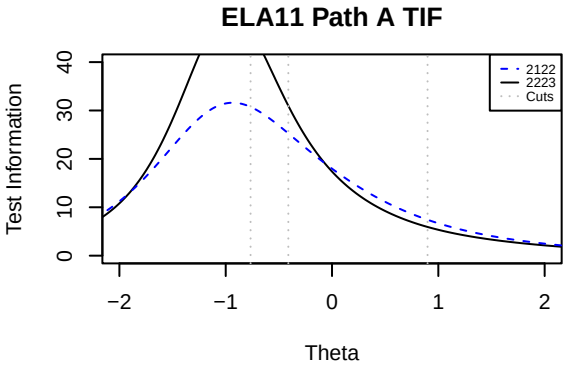

ELA08 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
30
50
−2
−1
0
1
2

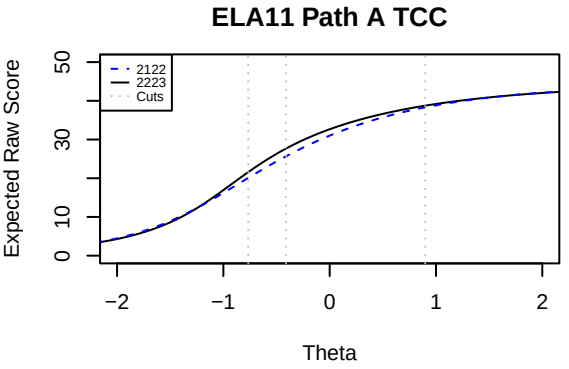

ELA08 Path B TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

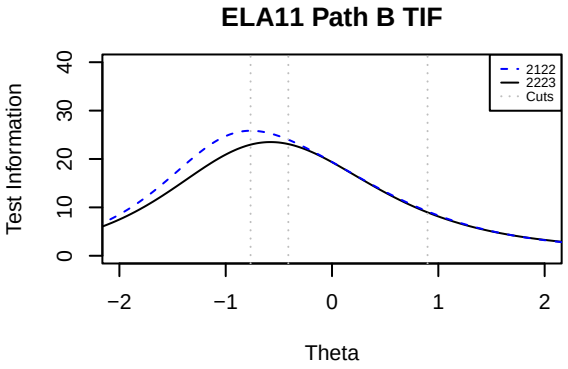

ELA08 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
30
50
−2
−1
0
1
2

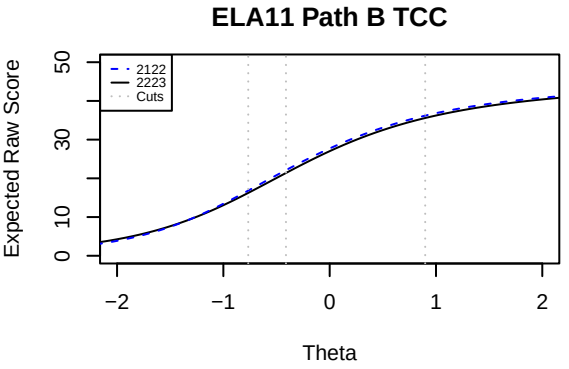

ELA08 Path C TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

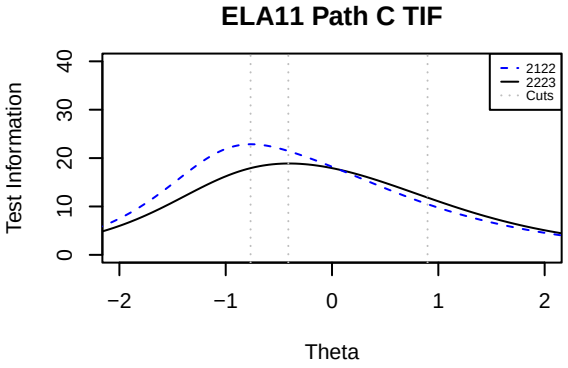

ELA08 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
30
50
−2
−1
0
1
2

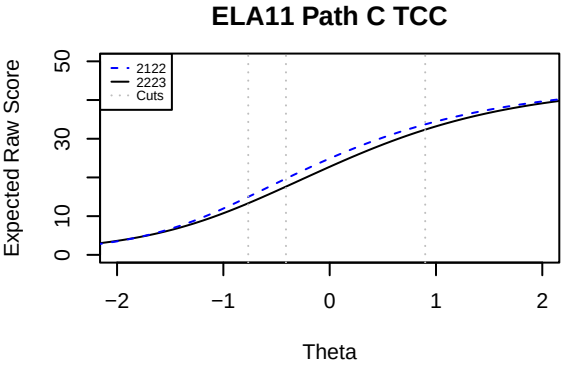

ELA11 Path A TIF
Test Information
Theta
2122
2223
Cuts

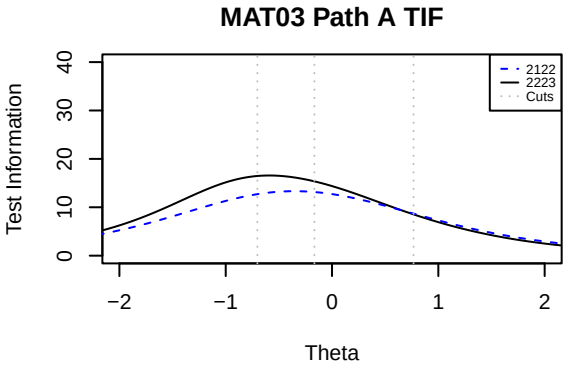

ELA11 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts

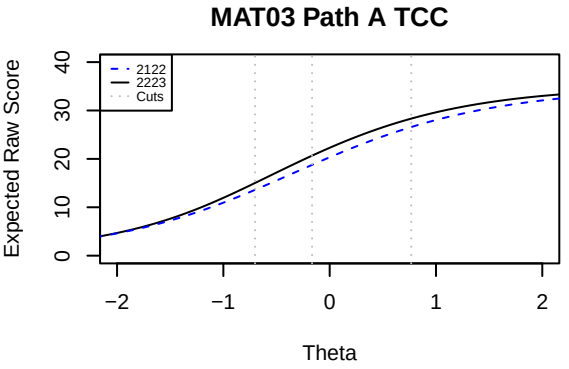

ELA11 Path B TIF
Test Information
Theta
2122
2223
Cuts

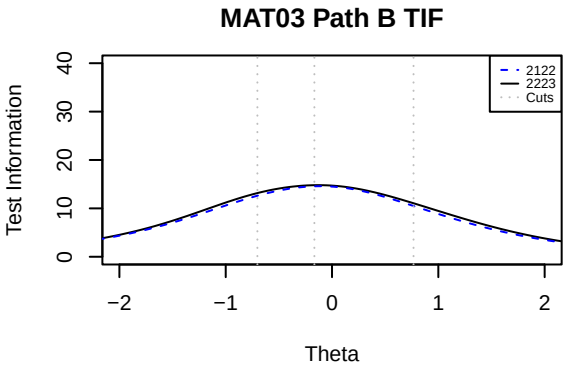

ELA11 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts

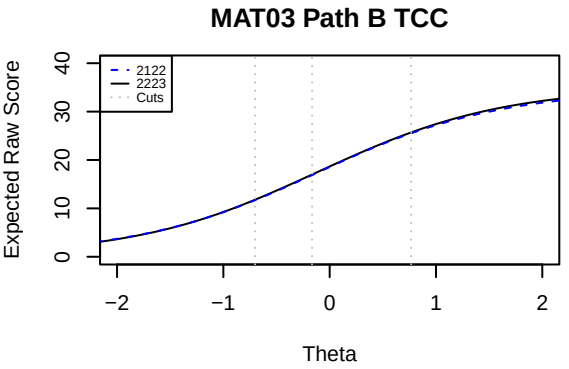

ELA11 Path C TIF
Test Information
Theta
2122
2223
Cuts

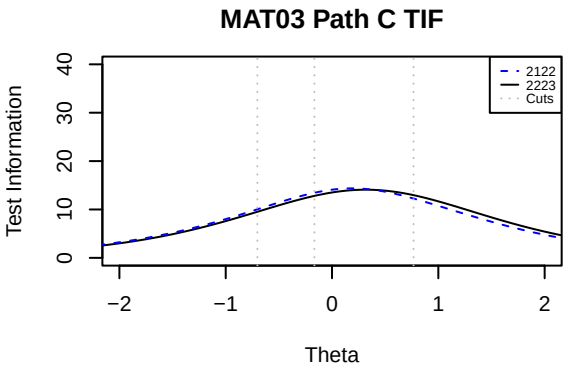

ELA11 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts

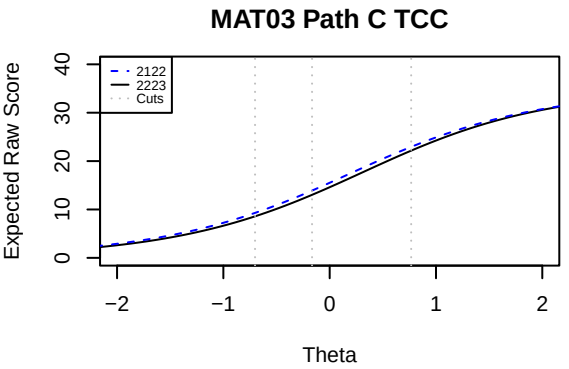

MAT03 Path A TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

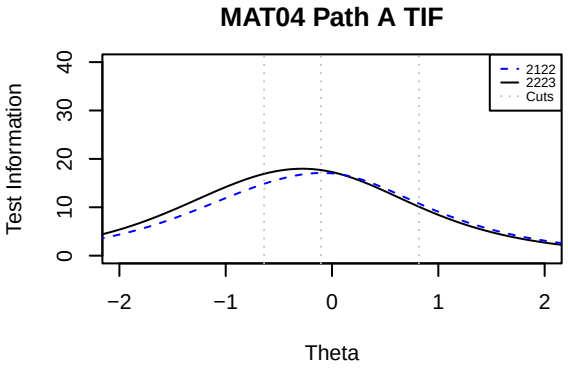

MAT03 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

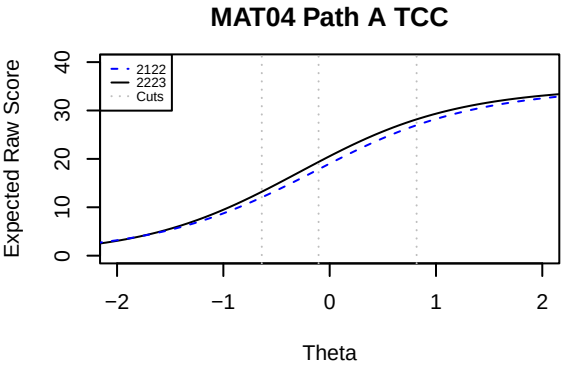

MAT03 Path B TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

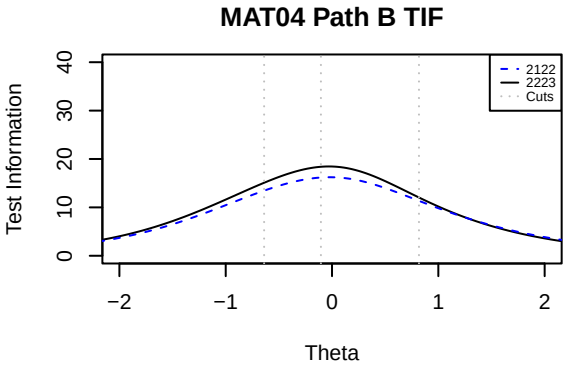

MAT03 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

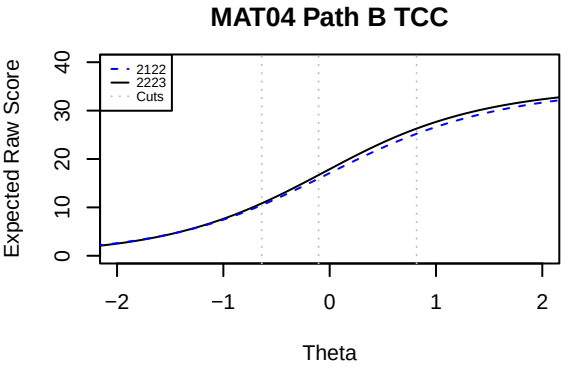

MAT03 Path C TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

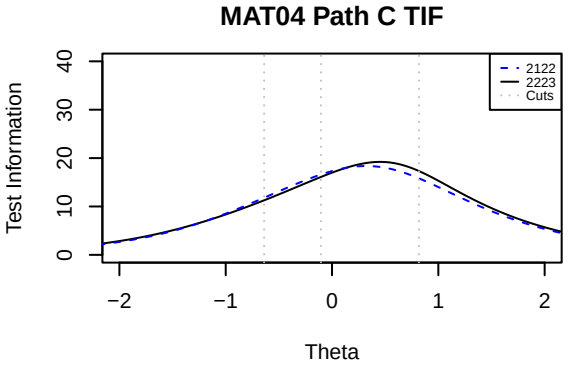

MAT03 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

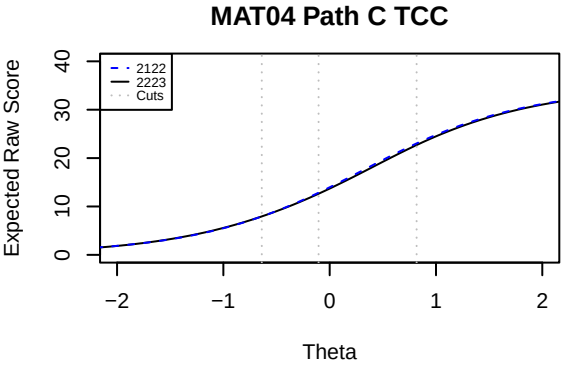

MAT04 Path A TIF
Test Information
Theta
2122
2223
Cuts

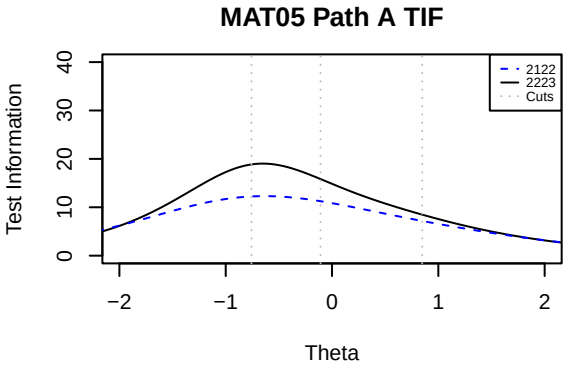

MAT04 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts

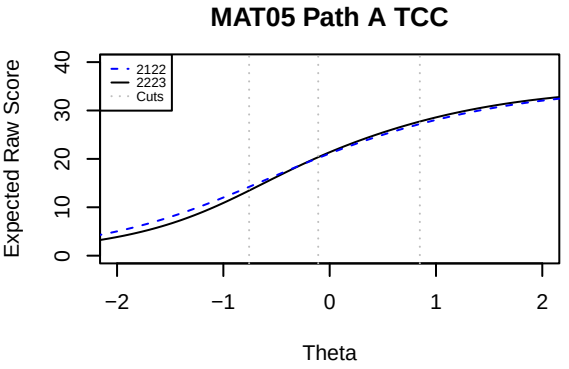

MAT04 Path B TIF
Test Information
Theta
2122
2223
Cuts

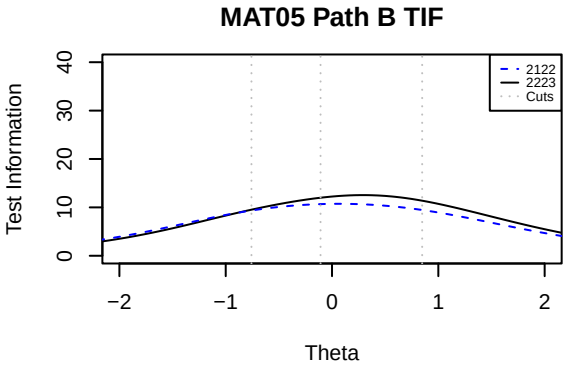

MAT04 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts

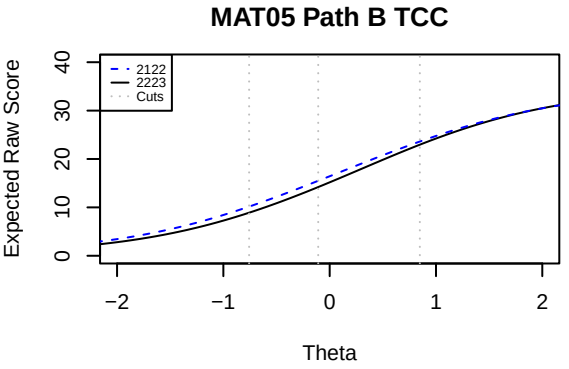

MAT04 Path C TIF
Test Information
Theta
2122
2223
Cuts

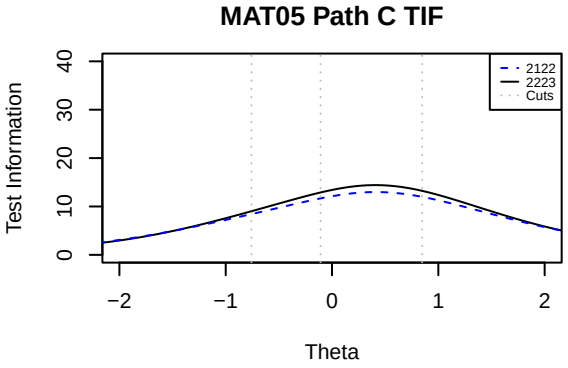

MAT04 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts

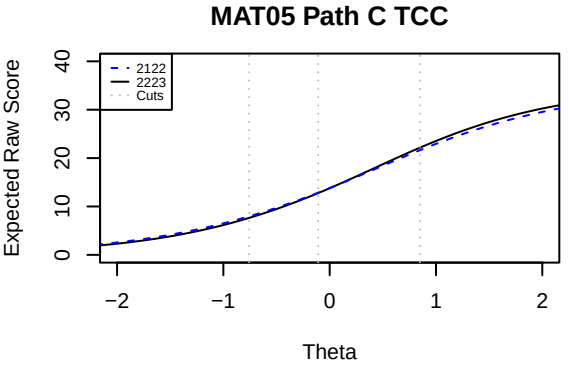

MAT05 Path A TIF
Test Information
Theta
2122
2223
Cuts

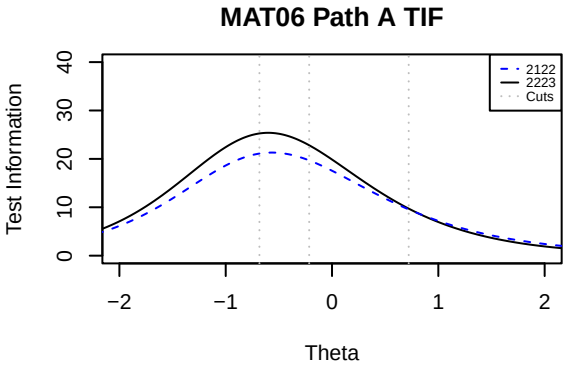

MAT05 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts

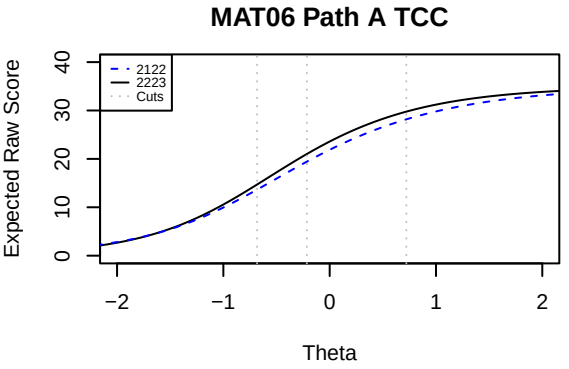

MAT05 Path B TIF
Test Information
Theta
2122
2223
Cuts

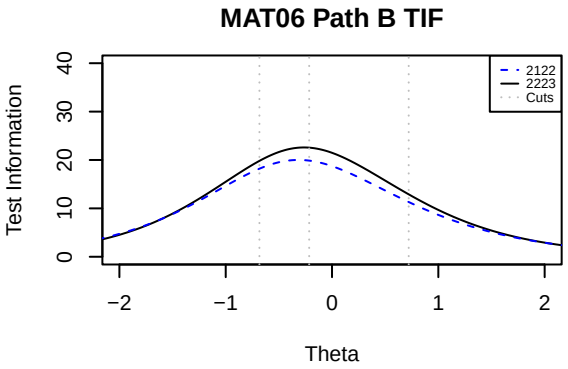

MAT05 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts

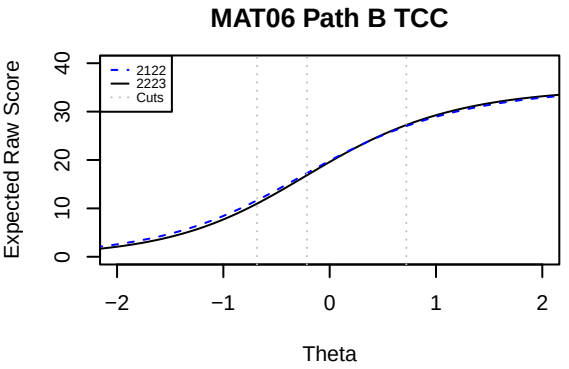

MAT05 Path C TIF
Test Information
Theta
2122
2223
Cuts

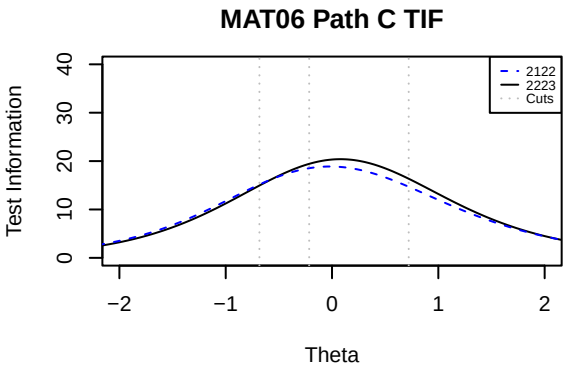

MAT05 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts

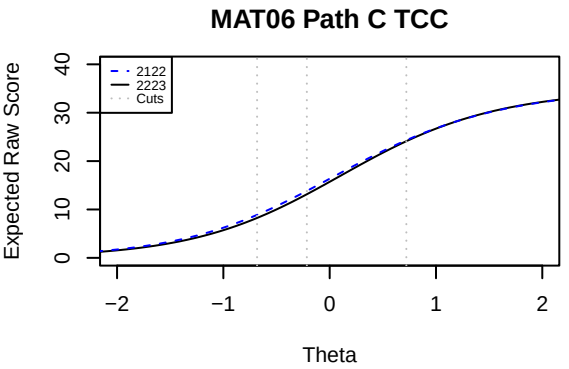

MAT06 Path A TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

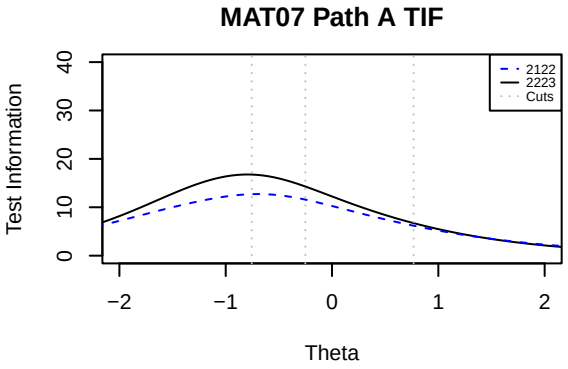

MAT06 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

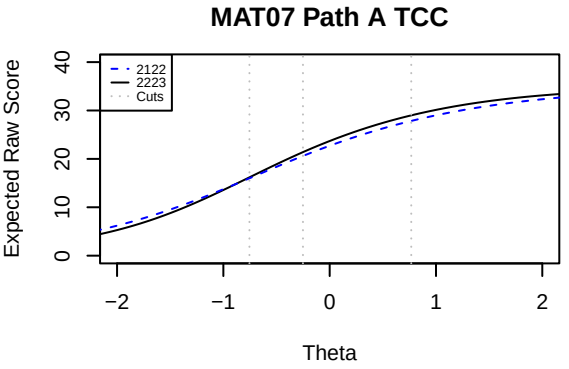

MAT06 Path B TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

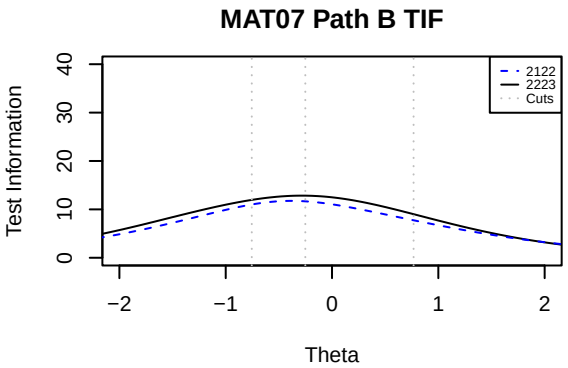

MAT06 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

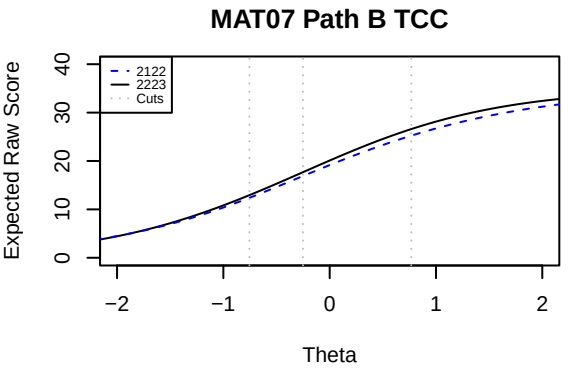

MAT06 Path C TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

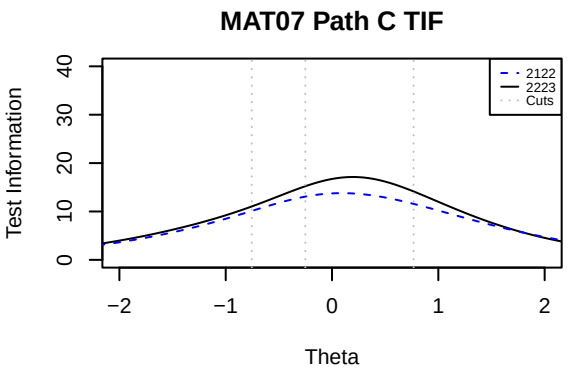

MAT06 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

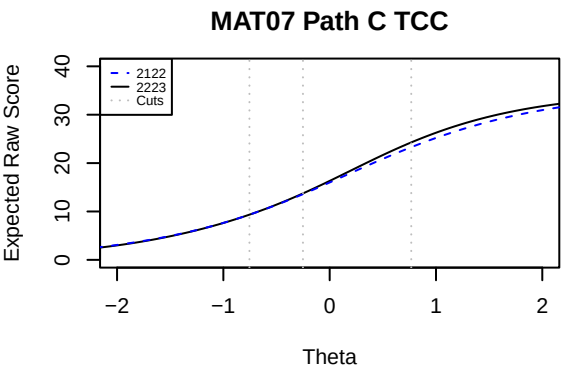

MAT07 Path A TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

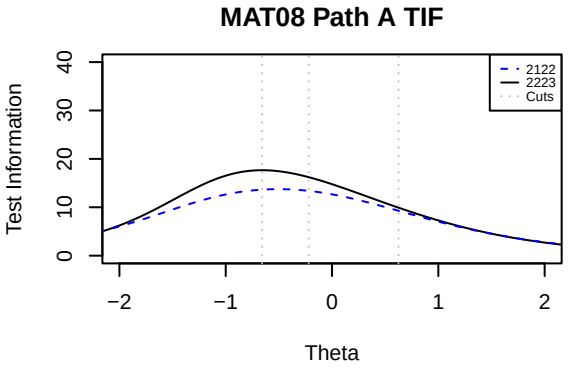

MAT07 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

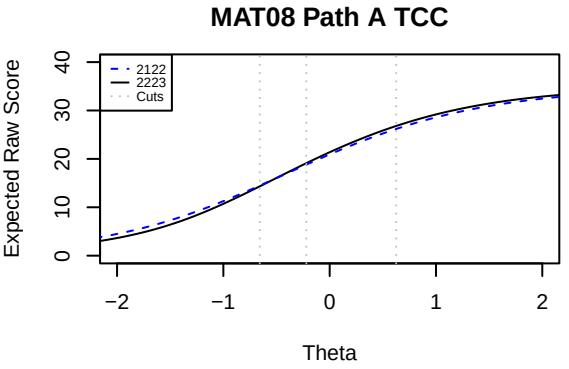

MAT07 Path B TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

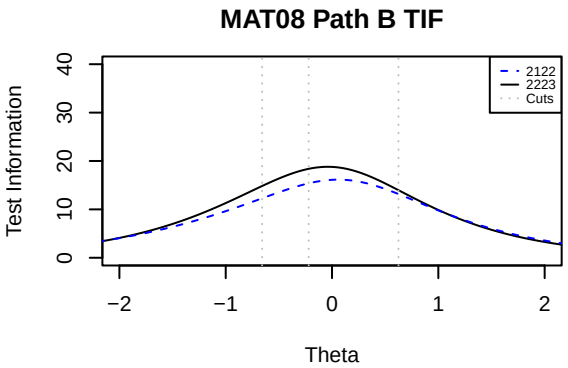

MAT07 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

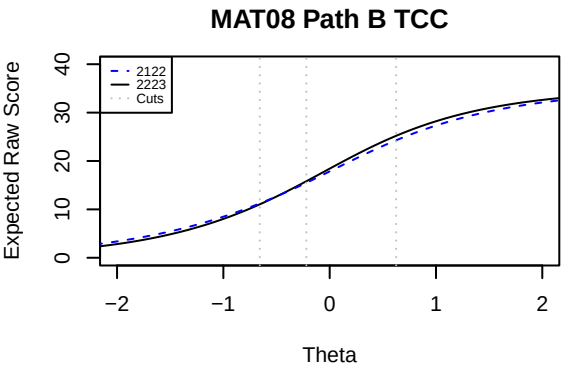

MAT07 Path C TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

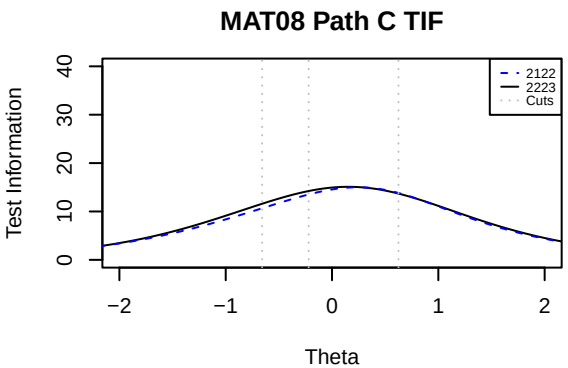

MAT07 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

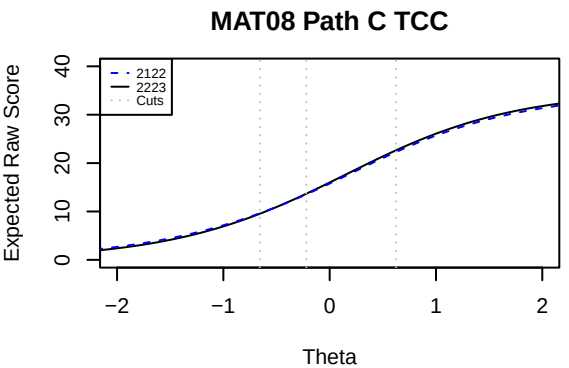

MAT08 Path A TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

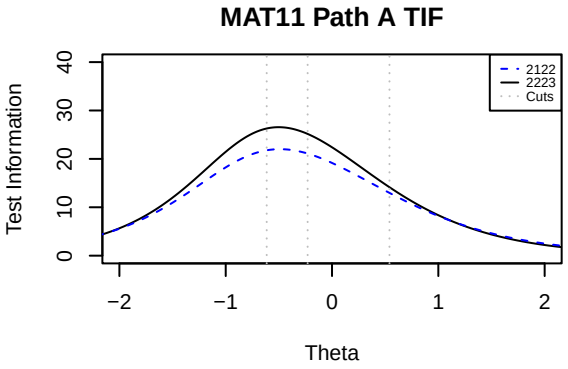

MAT08 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

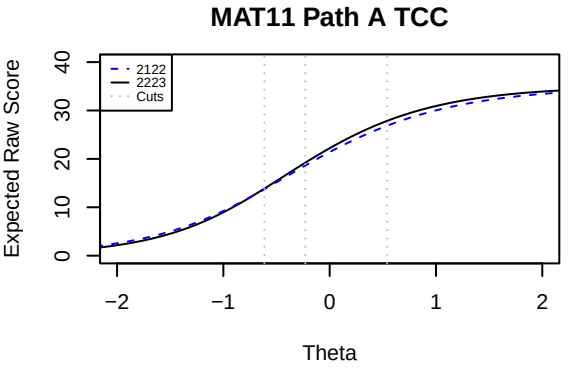

MAT08 Path B TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

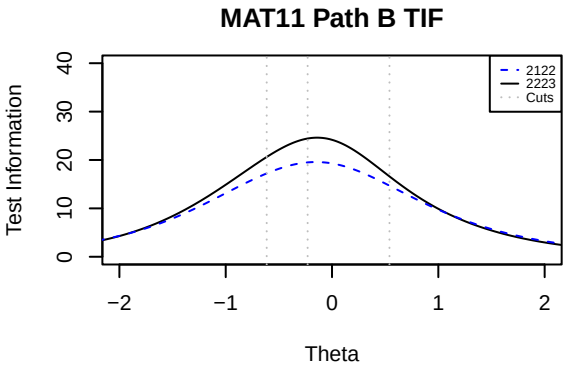

MAT08 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

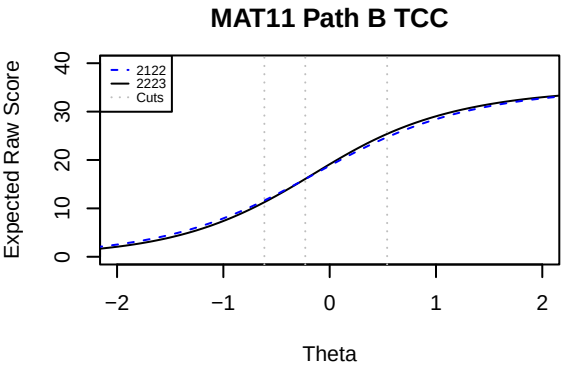

MAT08 Path C TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

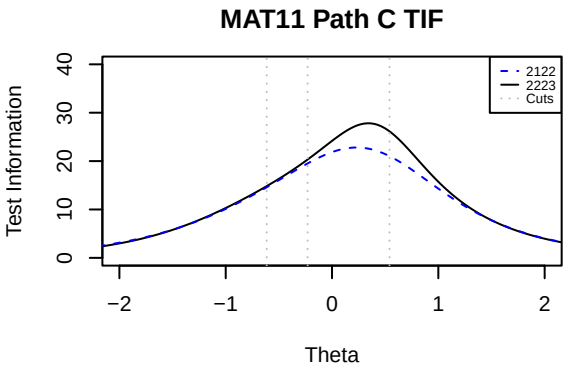

MAT08 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

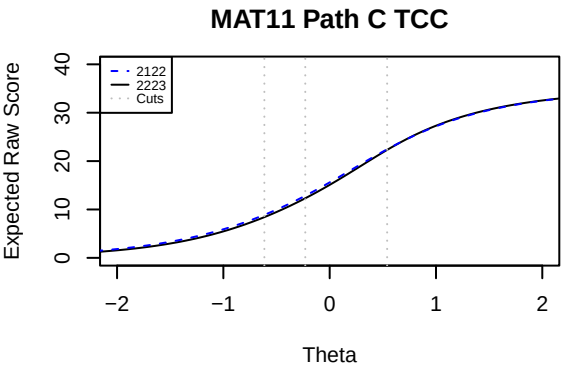

MAT11 Path A TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2


MAT11 Path A TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2


MAT11 Path B TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2


MAT11 Path B TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2


MAT11 Path C TIF
Test Information
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2


MAT11 Path C TCC
Expected Raw Score
Theta
2122
2223
Cuts
0
10
20
30
40
−2
−1
0
1
2

APPENDIX K
RAW TO SCALED SCORE LOOK-UP TABLES

Table K-1. Raw to Scaled Score Look-up Table—ELA Grade 3—2022 to 2023

Path	Raw Score	2022 Scaled Score	2022 Standard Error	2022 Performance Level	2023 Scaled Score	2023 Standard Error	2023 Performance Level
A	0	1200	16.92	1	1200	17.42	1
	1	1200	10.01	1	1200	13.03	1
	2	1206	6.64	1	1200	8.88	1
	3	1211	5.25	1	1206	6.61	1
	4	1213	4.47	1	1210	5.42	1
	5	1216	3.96	1	1212	4.68	1
	6	1217	3.59	1	1215	4.17	1
	7	1219	3.32	1	1217	3.79	1
	8	1220	3.11	1	1218	3.51	1
	9	1222	2.95	1	1220	3.28	1
	10	1223	2.81	1	1221	3.11	1
	11	1224	2.70	1	1222	2.96	1
	12	1225	2.61	1	1223	2.84	1
	13	1226	2.54	1	1224	2.75	1
	14	1227	2.48	1	1225	2.67	1
	15	1227	2.43	1	1226	2.60	1
	16	1228	2.39	1	1227	2.55	1
	17	1229	2.36	1	1228	2.51	1
	18	1230	2.33	1	1229	2.48	1
	19	1231	2.31	1	1230	2.46	1
	20	1231	2.30	1	1230	2.45	1
	21	1232	2.30	1	1231	2.45	1
	22	1233	2.30	1	1232	2.45	1
	23	1233	2.30	1	1233	2.47	1
	24	1234	2.31	2	1233	2.49	1
	25	1235	2.33	2	1234	2.52	2
	26	1236	2.35	2	1235	2.55	2
	27	1237	2.37	2	1236	2.60	2
	28	1238	2.41	2	1237	2.65	2
	29	1238	2.45	2	1238	2.72	2
	30	1239	2.50	2	1239	2.79	2
	31	1240	2.55	3	1240	2.88	3
	32	1241	2.62	3	1241	2.98	3
	33	1242	2.71	3	1242	3.09	3
	34	1243	2.80	3	1243	3.23	3
	35	1244	2.92	3	1245	3.38	3
	36	1245	3.05	3	1246	3.57	3
	37	1247	3.22	3	1248	3.79	3
	38	1248	3.41	3	1249	4.06	3
	39	1249	3.65	3	1251	4.39	3
	40	1251	3.96	3	1253	4.80	3
	41	1253	4.34	3	1256	5.33	4
	42	1256	4.85	4	1259	6.01	4
	43	1258	5.55	4	1262	6.96	4
	44	1262	6.58	4	1267	8.39	4
	45	1267	8.29	4	1275	10.99	4
	46	1276	12.11	4	1288	18.06	4
	47	1290	20.55	4	1290	18.61	4

continued

Path	Raw Score	2022 Scaled Score	2022 Standard Error	2022 Performance Level	2023 Scaled Score	2023 Standard Error	2023 Performance Level
B	0	1200	19.01	1	1200	19.24	1
	1	1200	11.36	1	1200	14.93	1
	2	1208	7.68	1	1200	11.02	1
	3	1212	6.12	1	1205	8.26	1
	4	1216	5.23	1	1210	6.75	1
	5	1218	4.64	1	1213	5.80	1
	6	1221	4.22	1	1216	5.15	1
	7	1222	3.90	1	1218	4.67	1
	8	1224	3.65	1	1220	4.30	1
	9	1226	3.46	1	1222	4.02	1
	10	1227	3.30	1	1224	3.79	1
	11	1228	3.16	1	1225	3.61	1
	12	1229	3.05	1	1227	3.46	1
	13	1230	2.96	1	1228	3.34	1
	14	1232	2.89	1	1229	3.24	1
	15	1233	2.82	1	1230	3.15	1
	16	1233	2.77	1	1231	3.09	1
	17	1234	2.73	2	1232	3.04	1
	18	1235	2.70	2	1233	3.00	1
	19	1236	2.68	2	1234	2.97	2
	20	1237	2.67	2	1235	2.96	2
	21	1238	2.66	2	1236	2.95	2
	22	1239	2.67	2	1237	2.95	2
	23	1240	2.69	3	1238	2.96	2
	24	1241	2.71	3	1239	2.98	2
	25	1242	2.74	3	1240	3.01	3
	26	1243	2.79	3	1241	3.05	3
	27	1244	2.84	3	1242	3.11	3
	28	1245	2.90	3	1244	3.17	3
	29	1246	2.96	3	1245	3.24	3
	30	1247	3.04	3	1246	3.34	3
	31	1248	3.13	3	1247	3.44	3
	32	1249	3.23	3	1248	3.57	3
	33	1250	3.34	3	1250	3.72	3
	34	1252	3.47	3	1251	3.90	3
	35	1253	3.61	3	1253	4.12	3
	36	1254	3.77	4	1255	4.36	4
	37	1256	3.96	4	1256	4.66	4
	38	1258	4.17	4	1259	5.00	4
	39	1260	4.42	4	1261	5.41	4
	40	1262	4.72	4	1264	5.88	4
	41	1264	5.09	4	1267	6.42	4
	42	1267	5.56	4	1271	7.04	4
	43	1271	6.17	4	1275	7.76	4
	44	1275	7.00	4	1280	8.75	4
	45	1280	8.29	4	1288	10.60	4
	46	1289	11.06	4	1289	11.04	4
	47	1290	11.06	4	1290	11.04	4

continued

Path	Raw Score	2022 Scaled Score	2022 Standard Error	2022 Performance Level	2023 Scaled Score	2023 Standard Error	2023 Performance Level
C	0	1200	17.97	1	1200	19.87	1
	1	1200	11.96	1	1200	14.71	1
	2	1205	8.00	1	1201	10.09	1
	3	1210	6.33	1	1208	7.62	1
	4	1214	5.37	1	1212	6.27	1
	5	1217	4.75	1	1215	5.42	1
	6	1219	4.30	1	1218	4.84	1
	7	1221	3.97	1	1220	4.41	1
	8	1222	3.70	1	1222	4.08	1
	9	1224	3.50	1	1224	3.82	1
	10	1225	3.33	1	1225	3.62	1
	11	1227	3.19	1	1226	3.45	1
	12	1228	3.07	1	1228	3.32	1
	13	1229	2.97	1	1229	3.21	1
	14	1230	2.89	1	1230	3.12	1
	15	1231	2.82	1	1231	3.05	1
	16	1232	2.76	1	1232	2.99	1
	17	1233	2.71	1	1233	2.95	1
	18	1233	2.67	1	1234	2.91	2
	19	1235	2.64	2	1235	2.89	2
	20	1235	2.62	2	1236	2.87	2
	21	1236	2.61	2	1237	2.87	2
	22	1237	2.60	2	1238	2.87	2
	23	1238	2.61	2	1239	2.89	2
	24	1239	2.62	2	1240	2.91	3
	25	1240	2.64	3	1241	2.94	3
	26	1241	2.67	3	1242	2.98	3
	27	1242	2.71	3	1243	3.03	3
	28	1242	2.76	3	1244	3.09	3
	29	1243	2.83	3	1245	3.17	3
	30	1244	2.90	3	1246	3.25	3
	31	1246	2.99	3	1248	3.36	3
	32	1247	3.10	3	1249	3.48	3
	33	1248	3.23	3	1250	3.62	3
	34	1249	3.38	3	1252	3.78	3
	35	1250	3.55	3	1253	3.97	3
	36	1252	3.76	3	1255	4.18	4
	37	1254	4.02	4	1257	4.43	4
	38	1255	4.32	4	1259	4.72	4
	39	1257	4.69	4	1261	5.05	4
	40	1260	5.13	4	1264	5.44	4
	41	1263	5.67	4	1267	5.90	4
	42	1266	6.30	4	1270	6.47	4
	43	1270	7.03	4	1274	7.16	4
	44	1275	7.89	4	1279	8.12	4
	45	1281	9.13	4	1286	9.80	4
	46	1289	11.47	4	1289	10.9	4
	47	1290	11.47	4	1290	10.9	4

Table K-2. Raw to Scaled Score Look-up Table—ELA Grade 4—2022 to 2023

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
A	0	1200	18.51	1	1200	17.76	1
	1	1200	11.66	1	1200	12.05	1
	2	1203	7.49	1	1201	7.74	1
	3	1207	5.79	1	1206	5.98	1
	4	1211	4.83	1	1210	4.99	1
	5	1213	4.21	1	1212	4.35	1
	6	1215	3.77	1	1214	3.90	1
	7	1217	3.45	1	1216	3.56	1
	8	1218	3.20	1	1217	3.30	1
	9	1220	3.00	1	1219	3.10	1
	10	1221	2.85	1	1220	2.93	1
	11	1222	2.72	1	1221	2.80	1
	12	1223	2.62	1	1222	2.69	1
	13	1224	2.54	1	1223	2.61	1
	14	1225	2.48	1	1224	2.54	1
	15	1225	2.43	1	1225	2.48	1
	16	1226	2.40	1	1225	2.44	1
	17	1227	2.37	1	1226	2.41	1
	18	1228	2.36	1	1227	2.39	1
	19	1229	2.35	1	1228	2.38	1
	20	1229	2.36	1	1229	2.38	1
	21	1230	2.37	1	1229	2.38	1
	22	1231	2.39	1	1230	2.40	1
	23	1232	2.42	1	1231	2.42	1
	24	1233	2.46	1	1232	2.45	1
	25	1233	2.50	1	1233	2.49	1
	26	1234	2.55	2	1233	2.54	1
	27	1235	2.61	2	1234	2.60	2
	28	1236	2.68	2	1235	2.67	2
	29	1237	2.76	2	1236	2.75	2
	30	1238	2.85	2	1237	2.83	2
	31	1239	2.94	2	1238	2.93	2
	32	1240	3.05	3	1239	3.05	2
	33	1241	3.17	3	1240	3.17	3
	34	1242	3.30	3	1241	3.32	3
	35	1244	3.45	3	1243	3.48	3
	36	1245	3.62	3	1244	3.66	3
	37	1247	3.81	3	1245	3.87	3
	38	1248	4.02	3	1247	4.11	3
	39	1250	4.27	3	1249	4.39	3
	40	1252	4.57	3	1251	4.72	3
	41	1254	4.93	3	1253	5.13	3
	42	1257	5.40	3	1256	5.65	3
	43	1260	6.04	4	1259	6.36	4
	44	1264	7.00	4	1263	7.43	4
	45	1269	8.68	4	1269	9.32	4
	46	1278	12.76	4	1280	14.12	4
	47	1290	19.25	4	1290	19.88	4

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
B	0	1200	20.93	1	1200	20.21	1
	1	1200	13.51	1	1200	12.25	1
	2	1203	8.75	1	1204	7.83	1
	3	1209	6.82	1	1209	6.05	1
	4	1213	5.74	1	1212	5.06	1
	5	1216	5.04	1	1215	4.43	1
	6	1218	4.54	1	1217	3.99	1
	7	1220	4.17	1	1219	3.67	1
	8	1222	3.89	1	1220	3.42	1
	9	1223	3.66	1	1222	3.23	1
	10	1225	3.47	1	1223	3.08	1
	11	1226	3.31	1	1224	2.97	1
	12	1227	3.18	1	1225	2.87	1
	13	1229	3.07	1	1226	2.79	1
	14	1230	2.98	1	1227	2.73	1
	15	1231	2.90	1	1228	2.68	1
	16	1232	2.84	1	1229	2.64	1
	17	1233	2.79	1	1230	2.61	1
	18	1233	2.75	1	1231	2.59	1
	19	1235	2.72	2	1232	2.58	1
	20	1236	2.70	2	1233	2.57	1
	21	1236	2.69	2	1233	2.56	1
	22	1237	2.69	2	1234	2.57	2
	23	1238	2.70	2	1235	2.58	2
	24	1239	2.72	2	1236	2.59	2
	25	1240	2.75	3	1237	2.62	2
	26	1241	2.79	3	1238	2.66	2
	27	1242	2.85	3	1239	2.70	2
	28	1243	2.91	3	1239	2.76	2
	29	1244	2.99	3	1241	2.83	3
	30	1245	3.07	3	1242	2.91	3
	31	1246	3.17	3	1243	3.01	3
	32	1247	3.29	3	1244	3.14	3
	33	1249	3.42	3	1245	3.28	3
	34	1250	3.57	3	1247	3.45	3
	35	1251	3.74	3	1248	3.64	3
	36	1253	3.93	3	1250	3.87	3
	37	1255	4.15	3	1251	4.14	3
	38	1256	4.42	3	1253	4.45	3
	39	1258	4.73	3	1255	4.82	3
	40	1261	5.10	4	1258	5.26	3
	41	1263	5.57	4	1261	5.78	4
	42	1266	6.15	4	1264	6.41	4
	43	1270	6.90	4	1268	7.20	4
	44	1275	7.93	4	1273	8.24	4
	45	1281	9.52	4	1279	9.78	4
	46	1288	12.03	4	1288	12.75	4
	47	1290	12.03	4	1290	12.75	4

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
C	0	1200	19.43	1	1200	19.29	1
	1	1200	12.91	1	1200	11.52	1
	2	1224	3.80	1	1223	3.21	1
	3	1226	3.67	1	1224	3.10	1
	4	1227	3.57	1	1225	3.03	1
	5	1228	3.49	1	1226	2.97	1
	6	1230	3.42	1	1227	2.93	1
	7	1231	3.36	1	1228	2.90	1
	8	1232	3.32	1	1229	2.89	1
	9	1233	3.29	1	1230	2.89	1
	10	1234	3.26	2	1231	2.90	1
	11	1235	3.24	2	1232	2.91	1
	12	1203	8.63	1	1204	7.59	1
	13	1236	3.23	2	1233	2.94	1
	14	1237	3.23	2	1234	2.97	2
	15	1239	3.23	2	1235	3.00	2
	16	1239	3.23	2	1236	3.04	2
	17	1241	3.24	3	1237	3.09	2
	18	1242	3.26	3	1239	3.13	2
	19	1243	3.28	3	1239	3.18	2
	20	1244	3.31	3	1241	3.24	3
	21	1245	3.34	3	1242	3.30	3
	22	1246	3.39	3	1243	3.36	3
	23	1208	6.86	1	1209	5.97	1
	24	1247	3.44	3	1244	3.43	3
	25	1249	3.51	3	1246	3.50	3
	26	1250	3.58	3	1247	3.59	3
	27	1251	3.67	3	1248	3.68	3
	28	1253	3.78	3	1250	3.79	3
	29	1254	3.91	3	1251	3.92	3
	30	1255	4.06	3	1253	4.07	3
	31	1257	4.24	3	1254	4.25	3
	32	1259	4.45	4	1256	4.47	3
	33	1261	4.72	4	1258	4.73	3
	34	1212	5.86	1	1212	5.06	1
	35	1263	5.04	4	1260	5.05	4
	36	1265	5.44	4	1263	5.45	4
	37	1268	5.95	4	1266	5.97	4
	38	1272	6.62	4	1269	6.65	4
	39	1276	7.57	4	1274	7.61	4
	40	1282	9.08	4	1280	9.10	4
	41	1288	11.28	4	1288	12.04	4
	42	1290	11.28	4	1290	12.04	4
	43	1215	5.21	1	1215	4.46	1
	44	1217	4.75	1	1217	4.05	1
	45	1219	4.42	1	1218	3.75	1
	46	1221	4.16	1	1220	3.52	1
	47	1223	3.96	1	1221	3.34	1

Table K-3. Raw to Scaled Score Look-up Table—ELA Grade 5—2022 to 2023

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
A	0	1200	21.09	1	1200	20.93	1
	1	1200	9.42	1	1200	10.41	1
	2	1207	6.07	1	1206	6.50	1
	3	1211	4.73	1	1210	4.98	1
	4	1214	3.97	1	1213	4.15	1
	5	1216	3.48	1	1215	3.62	1
	6	1217	3.14	1	1217	3.24	1
	7	1219	2.88	1	1218	2.97	1
	8	1220	2.68	1	1219	2.76	1
	9	1221	2.53	1	1221	2.59	1
	10	1222	2.40	1	1222	2.46	1
	11	1223	2.30	1	1223	2.35	1
	12	1224	2.22	1	1223	2.26	1
	13	1225	2.16	1	1224	2.19	1
	14	1225	2.11	1	1225	2.14	1
	15	1226	2.07	1	1226	2.10	1
	16	1227	2.05	1	1227	2.07	1
	17	1228	2.03	1	1227	2.05	1
	18	1228	2.03	1	1228	2.04	1
	19	1229	2.03	1	1229	2.04	1
	20	1230	2.05	1	1229	2.05	1
	21	1230	2.07	1	1230	2.07	1
	22	1231	2.10	1	1231	2.10	1
	23	1232	2.14	2	1232	2.13	2
	24	1233	2.19	2	1232	2.18	2
	25	1233	2.26	2	1233	2.23	2
	26	1234	2.33	2	1234	2.30	2
	27	1235	2.41	2	1235	2.37	2
	28	1236	2.50	2	1236	2.46	2
	29	1237	2.61	2	1237	2.56	2
	30	1238	2.73	2	1237	2.67	2
	31	1239	2.86	2	1239	2.79	2
	32	1240	3.00	3	1239	2.92	2
	33	1241	3.15	3	1241	3.07	3
	34	1243	3.32	3	1242	3.24	3
	35	1244	3.51	3	1243	3.42	3
	36	1246	3.73	3	1245	3.64	3
	37	1247	3.97	3	1247	3.89	3
	38	1249	4.25	3	1248	4.18	3
	39	1251	4.59	3	1251	4.54	3
	40	1254	5.02	3	1253	4.99	3
	41	1257	5.59	4	1256	5.60	4
	42	1260	6.42	4	1260	6.48	4
	43	1265	7.87	4	1265	7.98	4
	44	1274	11.41	4	1274	11.69	4
	45	1290	25.47	4	1290	26.12	4

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
B	0	1200	21.75	1	1200	21.64	1
	1	1200	12.93	1	1200	11.90	1
	2	1205	8.40	1	1206	7.36	1
	3	1210	6.56	1	1210	5.54	1
	4	1214	5.51	1	1213	4.54	1
	5	1217	4.83	1	1216	3.90	1
	6	1219	4.35	1	1218	3.46	1
	7	1221	4.00	1	1219	3.15	1
	8	1223	3.73	1	1221	2.91	1
	9	1224	3.53	1	1222	2.73	1
	10	1225	3.37	1	1223	2.60	1
	11	1227	3.25	1	1224	2.49	1
	12	1228	3.16	1	1225	2.42	1
	13	1229	3.08	1	1226	2.36	1
	14	1230	3.03	1	1227	2.33	1
	15	1232	2.99	2	1228	2.31	1
	16	1232	2.96	2	1229	2.31	1
	17	1233	2.94	2	1229	2.31	1
	18	1234	2.94	2	1230	2.34	1
	19	1235	2.94	2	1231	2.37	1
	20	1236	2.94	2	1232	2.41	2
	21	1237	2.96	2	1233	2.46	2
	22	1238	2.98	2	1234	2.52	2
	23	1239	3.01	2	1235	2.59	2
	24	1240	3.05	3	1236	2.67	2
	25	1241	3.10	3	1237	2.75	2
	26	1242	3.15	3	1238	2.84	2
	27	1244	3.21	3	1239	2.94	2
	28	1245	3.28	3	1239	3.04	2
	29	1246	3.36	3	1241	3.16	3
	30	1247	3.45	3	1242	3.28	3
	31	1248	3.56	3	1244	3.41	3
	32	1250	3.68	3	1245	3.56	3
	33	1251	3.82	3	1246	3.72	3
	34	1253	3.98	3	1248	3.91	3
	35	1254	4.17	3	1250	4.11	3
	36	1256	4.39	4	1252	4.35	3
	37	1258	4.65	4	1254	4.63	3
	38	1260	4.96	4	1255	4.96	3
	39	1263	5.34	4	1259	5.35	4
	40	1265	5.79	4	1261	5.83	4
	41	1269	6.36	4	1265	6.41	4
	42	1273	7.18	4	1269	7.15	4
	43	1279	8.61	4	1275	8.34	4
	44	1288	12.48	4	1284	11.36	4
	45	1290	14.46	4	1290	15.97	4

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
C	0	1200	22.31	1	1200	21.68	1
	1	1200	12.47	1	1200	12.06	1
	2	1206	8.23	1	1206	7.65	1
	3	1211	6.52	1	1211	5.81	1
	4	1215	5.55	1	1214	4.77	1
	5	1218	4.91	1	1217	4.11	1
	6	1220	4.46	1	1219	3.64	1
	7	1222	4.13	1	1220	3.31	1
	8	1224	3.88	1	1222	3.07	1
	9	1225	3.69	1	1223	2.89	1
	10	1227	3.54	1	1224	2.76	1
	11	1228	3.42	1	1225	2.66	1
	12	1229	3.33	1	1226	2.60	1
	13	1230	3.25	1	1227	2.57	1
	14	1232	3.20	2	1228	2.55	1
	15	1233	3.16	2	1229	2.56	1
	16	1234	3.13	2	1230	2.58	1
	17	1235	3.12	2	1232	2.62	2
	18	1236	3.11	2	1232	2.66	2
	19	1237	3.11	2	1233	2.72	2
	20	1238	3.11	2	1234	2.78	2
	21	1239	3.13	2	1235	2.85	2
	22	1240	3.15	3	1236	2.91	2
	23	1241	3.18	3	1237	2.98	2
	24	1242	3.21	3	1238	3.05	2
	25	1243	3.25	3	1239	3.12	2
	26	1245	3.30	3	1241	3.19	3
	27	1246	3.36	3	1242	3.26	3
	28	1247	3.42	3	1243	3.33	3
	29	1248	3.50	3	1244	3.41	3
	30	1249	3.59	3	1246	3.50	3
	31	1251	3.69	3	1247	3.60	3
	32	1252	3.81	3	1248	3.71	3
	33	1254	3.94	3	1250	3.83	3
	34	1255	4.10	3	1251	3.98	3
	35	1257	4.28	4	1253	4.15	3
	36	1259	4.49	4	1255	4.35	3
	37	1261	4.73	4	1257	4.59	4
	38	1263	5.02	4	1259	4.88	4
	39	1265	5.36	4	1261	5.22	4
	40	1268	5.77	4	1264	5.65	4
	41	1271	6.33	4	1267	6.18	4
	42	1276	7.15	4	1272	6.92	4
	43	1281	8.67	4	1277	8.16	4
	44	1290	12.73	4	1285	11.30	4
	45	1290	13.11	4	1290	14.74	4

Table K-4. Raw to Scaled Score Look-up Table—ELA Grade 6—2022 to 2023

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
	0	1200	27.50	1	1200	27.17	1
	1	1203	8.74	1	1203	8.46	1
	2	1209	5.73	1	1209	5.58	1
	3	1213	4.45	1	1212	4.38	1
	4	1215	3.71	1	1215	3.69	1
	5	1217	3.23	1	1217	3.23	1
	6	1219	2.88	1	1218	2.90	1
	7	1220	2.61	1	1220	2.64	1
	8	1221	2.40	1	1221	2.44	1
	9	1222	2.24	1	1222	2.28	1
	10	1223	2.10	1	1223	2.14	1
	11	1224	1.99	1	1223	2.02	1
	12	1225	1.89	1	1224	1.93	1
	13	1226	1.82	1	1225	1.85	1
	14	1226	1.75	1	1226	1.78	1
	15	1227	1.71	1	1226	1.73	1
	16	1227	1.67	1	1227	1.68	1
	17	1228	1.65	1	1227	1.65	1
	18	1229	1.64	1	1228	1.63	1
	19	1229	1.64	1	1229	1.62	1
	20	1230	1.65	1	1229	1.63	1
	21	1231	1.67	2	1230	1.64	1
A	22	1231	1.70	2	1231	1.66	2
	23	1232	1.74	2	1231	1.70	2
	24	1232	1.80	2	1232	1.74	2
	25	1233	1.86	2	1232	1.80	2
	26	1234	1.94	2	1233	1.87	2
	27	1234	2.03	2	1234	1.95	2
	28	1235	2.13	2	1234	2.05	2
	29	1236	2.24	2	1235	2.16	2
	30	1237	2.37	3	1236	2.29	2
	31	1238	2.52	3	1237	2.44	3
	32	1239	2.69	3	1238	2.61	3
	33	1240	2.88	3	1239	2.81	3
	34	1241	3.10	3	1241	3.03	3
	35	1243	3.34	3	1242	3.29	3
	36	1244	3.62	3	1244	3.58	3
	37	1246	3.95	3	1245	3.92	3
	38	1248	4.33	3	1247	4.31	3
	39	1251	4.79	4	1250	4.78	3
	40	1254	5.39	4	1253	5.37	4
	41	1257	6.23	4	1256	6.20	4
	42	1262	7.62	4	1261	7.56	4
	43	1270	10.80	4	1269	10.65	4
	44	1290	23.68	4	1290	24.90	4

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
	0	1200	26.35	1	1200	27.36	1
	1	1200	10.79	1	1200	10.82	1
	2	1208	7.24	1	1207	6.90	1
	3	1213	5.75	1	1212	5.38	1
	4	1216	4.88	1	1215	4.56	1
	5	1218	4.31	1	1217	4.03	1
	6	1220	3.89	1	1219	3.66	1
	7	1222	3.59	1	1221	3.38	1
	8	1224	3.35	1	1222	3.17	1
	9	1225	3.16	1	1223	3.00	1
	10	1226	3.02	1	1224	2.87	1
	11	1227	2.90	1	1226	2.76	1
	12	1228	2.81	1	1227	2.68	1
	13	1230	2.74	1	1228	2.61	1
	14	1231	2.68	2	1229	2.56	1
	15	1231	2.65	2	1230	2.52	1
	16	1232	2.62	2	1231	2.49	2
	17	1233	2.61	2	1231	2.48	2
	18	1234	2.62	2	1232	2.47	2
	19	1235	2.63	2	1233	2.48	2
	20	1236	2.66	2	1234	2.50	2
	21	1237	2.69	3	1235	2.53	2
B	22	1238	2.74	3	1236	2.57	2
	23	1239	2.79	3	1237	2.62	3
	24	1240	2.85	3	1238	2.68	3
	25	1241	2.92	3	1239	2.75	3
	26	1242	2.99	3	1240	2.83	3
	27	1243	3.08	3	1241	2.93	3
	28	1244	3.17	3	1242	3.03	3
	29	1245	3.27	3	1243	3.15	3
	30	1247	3.38	3	1244	3.29	3
	31	1248	3.51	3	1246	3.44	3
	32	1249	3.65	3	1247	3.61	3
	33	1251	3.81	4	1248	3.80	3
	34	1252	3.99	4	1250	4.03	3
	35	1254	4.20	4	1252	4.29	4
	36	1256	4.44	4	1254	4.59	4
	37	1258	4.74	4	1256	4.96	4
	38	1261	5.09	4	1259	5.41	4
	39	1263	5.54	4	1262	5.96	4
	40	1267	6.13	4	1266	6.67	4
	41	1271	6.96	4	1270	7.64	4
	42	1276	8.29	4	1277	9.19	4
	43	1285	11.35	4	1287	13.01	4
	44	1290	12.56	4	1290	13.29	4

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
	0	1200	27.71	1	1200	31.02	1
	1	1200	11.34	1	1202	10.48	1
	2	1208	7.38	1	1209	6.64	1
	3	1212	5.77	1	1214	5.18	1
	4	1216	4.86	1	1217	4.39	1
	5	1218	4.26	1	1219	3.88	1
	6	1220	3.83	1	1221	3.52	1
	7	1222	3.51	1	1222	3.26	1
	8	1223	3.27	1	1224	3.05	1
	9	1225	3.08	1	1225	2.89	1
	10	1226	2.93	1	1226	2.77	1
	11	1227	2.81	1	1227	2.66	1
	12	1228	2.71	1	1228	2.58	1
	13	1229	2.63	1	1229	2.51	1
	14	1231	2.57	2	1231	2.46	2
	15	1231	2.53	2	1231	2.42	2
	16	1232	2.50	2	1232	2.39	2
	17	1233	2.48	2	1233	2.38	2
	18	1234	2.48	2	1234	2.37	2
	19	1235	2.49	2	1235	2.37	2
	20	1235	2.51	2	1236	2.38	2
	21	1236	2.54	2	1236	2.40	2
C	22	1237	2.57	3	1237	2.42	3
	23	1238	2.62	3	1238	2.46	3
	24	1239	2.68	3	1239	2.51	3
	25	1240	2.75	3	1240	2.56	3
	26	1241	2.83	3	1241	2.63	3
	27	1242	2.92	3	1242	2.72	3
	28	1243	3.02	3	1243	2.81	3
	29	1244	3.14	3	1244	2.93	3
	30	1245	3.27	3	1245	3.06	3
	31	1247	3.41	3	1246	3.21	3
	32	1248	3.58	3	1248	3.38	3
	33	1250	3.77	3	1249	3.58	3
	34	1251	3.99	4	1251	3.82	4
	35	1253	4.24	4	1253	4.09	4
	36	1255	4.54	4	1254	4.40	4
	37	1257	4.90	4	1257	4.77	4
	38	1260	5.34	4	1259	5.21	4
	39	1263	5.88	4	1262	5.75	4
	40	1266	6.59	4	1266	6.44	4
	41	1271	7.56	4	1270	7.37	4
	42	1277	9.12	4	1276	8.85	4
	43	1287	12.98	4	1286	12.44	4
	44	1290	13.01	4	1290	13.10	4

Table K-5. Raw to Scaled Score Look-up Table—ELA Grade 7—2022 to 2023

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
	0	1200	21.93	1	1200	18.36	1
	1	1203	9.30	1	1200	9.69	1
	2	1210	6.21	1	1207	6.54	1
	3	1214	4.93	1	1211	5.19	1
	4	1216	4.19	1	1214	4.41	1
	5	1219	3.71	1	1216	3.89	1
	6	1220	3.37	1	1218	3.51	1
	7	1222	3.11	1	1220	3.23	1
	8	1223	2.91	1	1221	3.01	1
	9	1224	2.75	1	1222	2.84	1
	10	1225	2.63	1	1223	2.70	1
	11	1226	2.52	1	1224	2.58	1
	12	1227	2.44	1	1225	2.49	1
	13	1228	2.37	1	1226	2.41	1
	14	1229	2.32	1	1227	2.35	1
	15	1230	2.28	1	1228	2.30	1
	16	1231	2.25	1	1229	2.26	1
	17	1231	2.23	1	1229	2.23	1
	18	1232	2.22	1	1230	2.21	1
	19	1233	2.21	1	1231	2.20	1
	20	1234	2.22	1	1232	2.20	1
	21	1234	2.23	1	1232	2.21	1
A	22	1236	2.25	2	1233	2.22	1
	23	1236	2.28	2	1234	2.24	1
	24	1237	2.32	2	1235	2.27	1
	25	1238	2.36	2	1236	2.31	2
	26	1238	2.41	2	1236	2.36	2
	27	1239	2.48	2	1237	2.41	2
	28	1240	2.55	3	1238	2.48	2
	29	1241	2.63	3	1239	2.56	2
	30	1242	2.73	3	1239	2.65	2
	31	1243	2.85	3	1241	2.76	3
	32	1244	2.98	3	1242	2.88	3
	33	1246	3.13	3	1243	3.03	3
	34	1247	3.31	3	1244	3.21	3
	35	1248	3.52	3	1246	3.41	3
	36	1250	3.77	3	1247	3.66	3
	37	1252	4.07	3	1249	3.95	3
	38	1254	4.43	3	1251	4.32	3
	39	1256	4.88	4	1253	4.77	3
	40	1259	5.46	4	1256	5.35	4
	41	1262	6.30	4	1259	6.17	4
	42	1267	7.68	4	1264	7.46	4
	43	1275	10.92	4	1272	10.37	4
	44	1290	23.40	4	1290	27.35	4

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
	0	1200	24.76	1	1200	18.63	1
	1	1204	10.77	1	1200	9.78	1
	2	1212	7.29	1	1207	6.64	1
	3	1216	5.82	1	1211	5.30	1
	4	1219	4.98	1	1214	4.53	1
	5	1222	4.42	1	1217	4.03	1
	6	1224	4.03	1	1219	3.67	1
	7	1226	3.74	1	1220	3.40	1
	8	1227	3.51	1	1222	3.19	1
	9	1229	3.34	1	1223	3.02	1
	10	1230	3.19	1	1224	2.89	1
	11	1231	3.08	1	1225	2.79	1
	12	1232	2.99	1	1226	2.70	1
	13	1233	2.92	1	1227	2.64	1
	14	1234	2.86	1	1228	2.58	1
	15	1236	2.81	2	1229	2.54	1
	16	1236	2.78	2	1230	2.51	1
	17	1237	2.75	2	1231	2.49	1
	18	1238	2.74	2	1232	2.48	1
	19	1239	2.74	2	1233	2.48	1
	20	1240	2.74	3	1233	2.48	1
	21	1241	2.75	3	1234	2.50	1
B	22	1242	2.77	3	1236	2.52	2
	23	1243	2.80	3	1236	2.55	2
	24	1244	2.84	3	1237	2.59	2
	25	1245	2.89	3	1238	2.64	2
	26	1246	2.95	3	1239	2.69	2
	27	1247	3.02	3	1239	2.76	2
	28	1248	3.11	3	1241	2.84	3
	29	1249	3.20	3	1242	2.94	3
	30	1250	3.32	3	1243	3.05	3
	31	1252	3.45	3	1244	3.18	3
	32	1253	3.60	3	1246	3.33	3
	33	1255	3.77	4	1247	3.51	3
	34	1256	3.98	4	1248	3.72	3
	35	1258	4.23	4	1250	3.97	3
	36	1260	4.52	4	1252	4.28	3
	37	1262	4.86	4	1255	4.65	4
	38	1265	5.27	4	1257	5.11	4
	39	1268	5.75	4	1260	5.69	4
	40	1271	6.29	4	1264	6.38	4
	41	1276	6.94	4	1269	7.15	4
	42	1281	8.00	4	1275	7.95	4
	43	1290	11.08	4	1283	9.67	4
	44	1290	11.90	4	1290	13.73	4

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
C	0	1200	23.21	1	1200	19.91	1
	1	1203	10.45	1	1200	10.39	1
	2	1211	7.16	1	1208	7.01	1
	3	1215	5.76	1	1212	5.57	1
	4	1219	4.95	1	1216	4.74	1
	5	1221	4.41	1	1218	4.20	1
	6	1223	4.02	1	1220	3.81	1
	7	1225	3.73	1	1222	3.52	1
	8	1226	3.51	1	1223	3.30	1
	9	1228	3.33	1	1225	3.13	1
	10	1229	3.19	1	1226	3.00	1
	11	1230	3.07	1	1227	2.89	1
	12	1231	2.98	1	1228	2.81	1
	13	1232	2.90	1	1229	2.74	1
	14	1233	2.84	1	1230	2.68	1
	15	1234	2.79	1	1231	2.64	1
	16	1236	2.75	2	1232	2.61	1
	17	1236	2.73	2	1233	2.59	1
	18	1237	2.71	2	1234	2.58	1
	19	1238	2.70	2	1236	2.58	2
	20	1239	2.69	2	1236	2.58	2
	21	1239	2.70	2	1237	2.60	2
	22	1241	2.71	3	1238	2.62	2
	23	1242	2.74	3	1239	2.64	2
	24	1242	2.77	3	1239	2.68	2
	25	1243	2.81	3	1241	2.72	3
	26	1244	2.86	3	1242	2.78	3
	27	1245	2.92	3	1243	2.84	3
	28	1246	2.99	3	1244	2.92	3
	29	1247	3.07	3	1245	3.01	3
	30	1249	3.17	3	1246	3.12	3
	31	1250	3.29	3	1247	3.24	3
	32	1251	3.42	3	1249	3.39	3
	33	1252	3.58	3	1250	3.57	3
	34	1254	3.76	3	1252	3.78	3
	35	1256	3.98	4	1254	4.04	3
	36	1257	4.25	4	1255	4.35	4
	37	1259	4.57	4	1258	4.73	4
	38	1262	4.96	4	1260	5.21	4
	39	1265	5.44	4	1264	5.79	4
	40	1268	6.01	4	1267	6.46	4
	41	1272	6.68	4	1272	7.19	4
	42	1278	7.59	4	1278	8.11	4
	43	1286	9.76	4	1287	10.72	4
	44	1290	12.70	4	1290	12.92	4

Table K-6. Raw to Scaled Score Look-up Table—ELA Grade 8—2022 to 2023

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
	0	1200	24.61	1	1200	22.10	1
	1	1200	10.62	1	1200	10.90	1
	2	1207	6.51	1	1205	7.05	1
	3	1211	4.82	1	1209	5.40	1
	4	1214	3.91	1	1212	4.47	1
	5	1216	3.35	1	1215	3.86	1
	6	1217	2.97	1	1216	3.44	1
	7	1219	2.70	1	1218	3.13	1
	8	1220	2.51	1	1219	2.90	1
	9	1221	2.36	1	1221	2.73	1
	10	1222	2.25	1	1222	2.59	1
	11	1223	2.16	1	1223	2.49	1
	12	1223	2.09	1	1224	2.41	1
	13	1224	2.04	1	1224	2.34	1
	14	1225	2.00	1	1225	2.29	1
	15	1226	1.97	1	1226	2.25	1
	16	1226	1.95	1	1227	2.23	1
	17	1227	1.94	1	1228	2.21	1
	18	1228	1.94	1	1228	2.20	1
	19	1229	1.94	1	1229	2.21	1
	20	1229	1.94	1	1229	2.21	1
	21	1229	1.95	1	1231	2.23	2
A	22	1231	1.97	2	1231	2.26	2
	23	1231	1.99	2	1232	2.29	2
	24	1232	2.02	2	1233	2.34	2
	25	1233	2.06	2	1234	2.39	2
	26	1234	2.11	2	1235	2.45	2
	27	1234	2.19	2	1236	2.52	2
	28	1235	2.28	2	1237	2.61	2
	29	1236	2.40	2	1238	2.71	3
	30	1237	2.56	2	1239	2.83	3
	31	1238	2.74	3	1240	2.96	3
	32	1240	2.96	3	1241	3.12	3
	33	1241	3.22	3	1242	3.31	3
	34	1242	3.51	3	1244	3.52	3
	35	1244	3.84	3	1245	3.78	3
	36	1246	4.22	3	1247	4.10	3
	37	1248	4.65	3	1249	4.48	3
	38	1251	5.18	4	1251	4.96	4
	39	1253	5.82	4	1254	5.58	4
	40	1257	6.64	4	1257	6.40	4
	41	1262	7.76	4	1262	7.54	4
	42	1268	9.49	4	1268	9.31	4
	43	1278	13.10	4	1278	13.00	4
	44	1290	20.01	4	1290	20.00	4

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
B	0	1200	23.40	1	1200	23.82	1
	1	1200	11.89	1	1200	11.43	1
	2	1206	7.90	1	1206	7.42	1
	3	1211	6.21	1	1211	5.74	1
	4	1214	5.23	1	1214	4.79	1
	5	1217	4.59	1	1216	4.18	1
	6	1219	4.13	1	1218	3.76	1
	7	1221	3.79	1	1220	3.45	1
	8	1222	3.52	1	1222	3.21	1
	9	1224	3.32	1	1223	3.03	1
	10	1225	3.16	1	1224	2.89	1
	11	1226	3.03	1	1225	2.78	1
	12	1227	2.93	1	1226	2.69	1
	13	1228	2.86	1	1227	2.62	1
	14	1229	2.80	1	1228	2.56	1
	15	1230	2.76	2	1229	2.53	1
	16	1231	2.73	2	1229	2.50	1
	17	1232	2.71	2	1231	2.48	2
	18	1233	2.71	2	1232	2.47	2
	19	1234	2.71	2	1232	2.48	2
	20	1235	2.72	2	1233	2.49	2
	21	1236	2.74	2	1234	2.50	2
	22	1237	2.77	2	1235	2.53	2
	23	1238	2.81	3	1236	2.56	2
	24	1239	2.85	3	1237	2.61	2
	25	1240	2.90	3	1238	2.66	3
	26	1241	2.96	3	1239	2.71	3
	27	1242	3.02	3	1240	2.78	3
	28	1243	3.10	3	1241	2.86	3
	29	1244	3.18	3	1242	2.95	3
	30	1245	3.28	3	1243	3.06	3
	31	1246	3.39	3	1244	3.17	3
	32	1248	3.51	3	1245	3.31	3
	33	1249	3.65	3	1247	3.47	3
	34	1251	3.81	4	1248	3.64	3
	35	1252	3.98	4	1250	3.84	4
	36	1254	4.17	4	1252	4.07	4
	37	1256	4.37	4	1254	4.32	4
	38	1258	4.60	4	1256	4.59	4
	39	1261	4.87	4	1258	4.88	4
	40	1264	5.26	4	1261	5.24	4
	41	1267	5.93	4	1265	5.80	4
	42	1272	7.30	4	1270	6.96	4
	43	1280	11.20	4	1278	10.46	4
	44	1290	18.14	4	1290	20.18	4

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
C	0	1200	23.54	1	1200	23.61	1
	1	1200	11.43	1	1200	12.20	1
	2	1206	7.60	1	1205	7.89	1
	3	1211	5.98	1	1210	6.07	1
	4	1214	5.04	1	1214	5.04	1
	5	1217	4.43	1	1216	4.38	1
	6	1219	3.99	1	1218	3.92	1
	7	1221	3.67	1	1220	3.58	1
	8	1222	3.42	1	1222	3.33	1
	9	1223	3.22	1	1223	3.14	1
	10	1225	3.07	1	1224	2.99	1
	11	1226	2.94	1	1225	2.87	1
	12	1227	2.85	1	1226	2.78	1
	13	1228	2.77	1	1228	2.71	1
	14	1229	2.71	1	1229	2.66	1
	15	1229	2.67	1	1229	2.62	1
	16	1231	2.64	2	1230	2.60	2
	17	1232	2.62	2	1231	2.58	2
	18	1233	2.61	2	1232	2.58	2
	19	1233	2.61	2	1233	2.59	2
	20	1234	2.62	2	1234	2.61	2
	21	1235	2.64	2	1235	2.63	2
	22	1236	2.66	2	1236	2.66	2
	23	1237	2.69	2	1237	2.70	2
	24	1238	2.74	3	1238	2.75	3
	25	1239	2.78	3	1239	2.81	3
	26	1240	2.84	3	1240	2.87	3
	27	1241	2.91	3	1241	2.94	3
	28	1242	2.99	3	1242	3.03	3
	29	1243	3.07	3	1243	3.12	3
	30	1244	3.17	3	1244	3.22	3
	31	1245	3.29	3	1245	3.34	3
	32	1247	3.42	3	1247	3.47	3
	33	1248	3.57	3	1248	3.62	3
	34	1250	3.73	4	1250	3.79	4
	35	1251	3.92	4	1251	3.97	4
	36	1253	4.12	4	1253	4.18	4
	37	1255	4.35	4	1255	4.40	4
	38	1257	4.59	4	1258	4.64	4
	39	1260	4.86	4	1260	4.92	4
	40	1263	5.23	4	1263	5.31	4
	41	1266	5.83	4	1267	5.97	4
	42	1271	7.07	4	1272	7.35	4
	43	1279	10.68	4	1280	11.43	4
	44	1290	19.10	4	1290	18.65	4

Table K-7. Raw to Scaled Score Look-up Table—ELA Grade HS—2022 to 2023

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
	0	1200	22.42	1	1200	23.97	1
	1	1207	9.90	1	1209	8.99	1
	2	1214	6.10	1	1216	5.61	1
	3	1218	4.54	1	1219	4.25	1
	4	1221	3.68	1	1222	3.51	1
	5	1223	3.13	1	1223	3.04	1
	6	1224	2.75	1	1225	2.71	1
	7	1225	2.48	1	1226	2.48	1
	8	1226	2.27	1	1227	2.31	1
	9	1227	2.11	1	1228	2.17	1
	10	1228	1.98	1	1229	2.07	1
	11	1229	1.88	1	1230	1.98	1
	12	1230	1.80	1	1230	1.92	1
	13	1230	1.74	1	1231	1.87	1
	14	1231	1.70	1	1232	1.83	1
	15	1232	1.66	1	1232	1.80	1
	16	1232	1.64	1	1233	1.78	1
	17	1233	1.63	1	1234	1.77	1
	18	1233	1.62	1	1234	1.77	1
	19	1234	1.63	1	1235	1.78	1
	20	1234	1.64	1	1235	1.79	1
	21	1235	1.67	1	1236	1.81	2
A	22	1236	1.70	2	1237	1.84	2
	23	1236	1.74	2	1237	1.87	2
	24	1237	1.79	2	1238	1.91	2
	25	1238	1.85	2	1239	1.96	2
	26	1238	1.92	2	1239	2.02	2
	27	1239	2.00	2	1240	2.08	3
	28	1240	2.10	3	1241	2.16	3
	29	1241	2.21	3	1242	2.25	3
	30	1241	2.33	3	1242	2.34	3
	31	1242	2.47	3	1243	2.46	3
	32	1243	2.64	3	1244	2.59	3
	33	1245	2.82	3	1245	2.74	3
	34	1246	3.04	3	1247	2.92	3
	35	1247	3.28	3	1248	3.13	3
	36	1249	3.57	3	1249	3.38	3
	37	1251	3.92	3	1251	3.68	3
	38	1253	4.34	3	1253	4.06	3
	39	1255	4.86	4	1255	4.54	4
	40	1258	5.53	4	1258	5.16	4
	41	1262	6.43	4	1261	6.04	4
	42	1267	7.83	4	1266	7.40	4
	43	1275	10.90	4	1274	10.36	4
	44	1290	21.48	4	1290	22.37	4

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
	0	1200	21.35	1	1200	22.29	1
	1	1206	10.00	1	1208	9.61	1
	2	1214	6.56	1	1215	6.26	1
	3	1218	5.14	1	1219	4.84	1
	4	1221	4.35	1	1221	4.03	1
	5	1223	3.84	1	1223	3.50	1
	6	1225	3.49	1	1225	3.13	1
	7	1226	3.23	1	1227	2.86	1
	8	1227	3.03	1	1228	2.66	1
	9	1229	2.87	1	1229	2.51	1
	10	1230	2.75	1	1230	2.39	1
	11	1231	2.66	1	1231	2.30	1
	12	1232	2.58	1	1232	2.23	1
	13	1233	2.52	1	1232	2.19	1
	14	1234	2.47	1	1233	2.15	1
	15	1234	2.43	1	1234	2.13	1
	16	1235	2.40	1	1235	2.12	1
	17	1236	2.38	2	1235	2.11	1
	18	1237	2.37	2	1236	2.12	2
	19	1238	2.37	2	1237	2.14	2
	20	1238	2.37	2	1238	2.16	2
	21	1239	2.39	2	1238	2.18	2
B	22	1240	2.40	3	1239	2.22	2
	23	1241	2.43	3	1240	2.26	3
	24	1242	2.46	3	1241	2.31	3
	25	1242	2.50	3	1241	2.36	3
	26	1243	2.55	3	1242	2.42	3
	27	1244	2.61	3	1243	2.49	3
	28	1245	2.68	3	1244	2.56	3
	29	1246	2.76	3	1245	2.65	3
	30	1247	2.86	3	1246	2.75	3
	31	1248	2.97	3	1247	2.87	3
	32	1249	3.11	3	1248	3.00	3
	33	1251	3.27	3	1249	3.15	3
	34	1252	3.45	3	1251	3.33	3
	35	1254	3.68	3	1252	3.55	3
	36	1255	3.96	4	1254	3.81	3
	37	1257	4.30	4	1256	4.14	4
	38	1260	4.73	4	1258	4.55	4
	39	1262	5.29	4	1261	5.10	4
	40	1266	6.01	4	1264	5.83	4
	41	1270	6.96	4	1268	6.84	4
	42	1276	8.31	4	1274	8.25	4
	43	1285	11.04	4	1283	10.94	4
	44	1290	13.56	4	1290	14.14	4

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
C	0	1200	23.22	1	1200	24.36	1
	1	1208	10.24	1	1209	9.58	1
	2	1215	6.68	1	1216	6.15	1
	3	1219	5.23	1	1220	4.72	1
	4	1222	4.44	1	1223	3.91	1
	5	1225	3.93	1	1225	3.40	1
	6	1226	3.59	1	1226	3.05	1
	7	1228	3.34	1	1228	2.80	1
	8	1229	3.16	1	1229	2.62	1
	9	1231	3.02	1	1230	2.49	1
	10	1232	2.92	1	1231	2.39	1
	11	1233	2.83	1	1232	2.32	1
	12	1234	2.77	1	1233	2.27	1
	13	1235	2.73	1	1234	2.24	1
	14	1236	2.69	2	1235	2.23	1
	15	1237	2.67	2	1236	2.23	2
	16	1238	2.65	2	1236	2.24	2
	17	1239	2.65	2	1237	2.25	2
	18	1240	2.64	3	1238	2.28	2
	19	1241	2.65	3	1239	2.31	2
	20	1242	2.66	3	1240	2.35	3
	21	1243	2.67	3	1240	2.39	3
	22	1244	2.70	3	1241	2.44	3
	23	1244	2.72	3	1242	2.49	3
	24	1245	2.75	3	1243	2.55	3
	25	1246	2.79	3	1244	2.61	3
	26	1247	2.83	3	1245	2.68	3
	27	1248	2.89	3	1246	2.75	3
	28	1249	2.95	3	1247	2.83	3
	29	1251	3.01	3	1248	2.92	3
	30	1252	3.10	3	1249	3.01	3
	31	1253	3.19	3	1250	3.12	3
	32	1254	3.30	3	1252	3.25	3
	33	1255	3.43	4	1253	3.39	3
	34	1257	3.58	4	1254	3.56	3
	35	1258	3.76	4	1256	3.75	4
	36	1260	3.98	4	1258	3.98	4
	37	1262	4.25	4	1260	4.26	4
	38	1264	4.58	4	1262	4.61	4
	39	1267	5.02	4	1265	5.05	4
	40	1270	5.60	4	1268	5.64	4
	41	1274	6.42	4	1272	6.47	4
	42	1279	7.72	4	1277	7.76	4
	43	1287	10.58	4	1286	10.53	4
	44	1290	12.10	4	1290	12.75	4

Table K-8. Raw to Scaled Score Look-up Table—Mathematics Grade 3—2022 to 2023

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
A	0	1200	21.39	1	1200	23.37	1
	1	1200	12.24	1	1200	12.53	1
	2	1206	8.40	1	1208	8.46	1
	3	1212	6.71	1	1213	6.68	1
	4	1215	5.71	1	1217	5.64	1
	5	1218	5.04	1	1220	4.95	1
	6	1221	4.56	1	1222	4.48	1
	7	1223	4.20	1	1224	4.13	1
	8	1225	3.93	1	1226	3.87	1
	9	1226	3.72	1	1228	3.68	1
	10	1228	3.56	1	1229	3.54	1
	11	1229	3.44	1	1231	3.44	1
	12	1231	3.35	1	1232	3.37	1
	13	1232	3.29	1	1234	3.33	1
	14	1233	3.24	1	1235	3.30	2
	15	1234	3.22	1	1236	3.30	2
	16	1236	3.21	2	1237	3.31	2
	17	1237	3.21	2	1239	3.33	2
	18	1238	3.23	2	1240	3.37	2
	19	1239	3.26	2	1241	3.41	2
	20	1241	3.30	2	1243	3.47	3
	21	1242	3.35	3	1244	3.55	3
	22	1243	3.42	3	1245	3.63	3
	23	1245	3.51	3	1247	3.74	3
	24	1246	3.61	3	1249	3.87	3
	25	1248	3.74	3	1250	4.02	3
	26	1249	3.91	3	1252	4.21	3
	27	1251	4.11	3	1254	4.44	4
	28	1253	4.37	3	1256	4.74	4
	29	1255	4.70	4	1258	5.12	4
	30	1258	5.14	4	1261	5.63	4
	31	1261	5.75	4	1264	6.35	4
	32	1264	6.66	4	1269	7.43	4
	33	1270	8.23	4	1275	9.28	4
	34	1278	11.82	4	1284	13.50	4
	35	1290	24.68	4	1290	20.62	4
B	0	1200	29.01	1	1200	24.06	1
	1	1205	12.47	1	1200	12.82	1
	2	1214	8.65	1	1209	8.84	1
	3	1219	7.01	1	1214	7.11	1
	4	1223	6.05	1	1218	6.10	1
	5	1227	5.42	1	1221	5.41	1
	6	1229	4.97	1	1224	4.92	1
	7	1231	4.62	1	1226	4.56	1
	8	1233	4.36	1	1228	4.28	1
	9	1235	4.15	2	1230	4.07	1
	10	1237	3.98	2	1232	3.90	1
	11	1239	3.84	2	1233	3.78	1
	12	1240	3.74	2	1235	3.69	2
	13	1241	3.65	2	1236	3.62	2
	14	1243	3.58	3	1238	3.57	2
	15	1244	3.53	3	1239	3.53	2
	16	1246	3.50	3	1240	3.52	2
	17	1247	3.48	3	1242	3.51	3
	18	1248	3.48	3	1243	3.52	3
	19	1249	3.49	3	1244	3.54	3

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
B	20	1251	3.52	3	1246	3.58	3
	21	1252	3.56	3	1247	3.64	3
	22	1253	3.62	3	1249	3.71	3
	23	1255	3.69	4	1250	3.81	3
	24	1256	3.79	4	1252	3.93	3
	25	1258	3.92	4	1253	4.07	3
	26	1260	4.08	4	1255	4.25	4
	27	1261	4.27	4	1257	4.48	4
	28	1263	4.52	4	1259	4.76	4
	29	1266	4.85	4	1261	5.12	4
	30	1268	5.28	4	1264	5.61	4
	31	1271	5.88	4	1267	6.29	4
	32	1275	6.78	4	1271	7.33	4
	33	1280	8.33	4	1277	9.12	4
	34	1289	11.95	4	1287	13.25	4
	35	1290	16.00	4	1290	18.87	4
C	0	1200	24.16	1	1200	27.67	1
	1	1201	12.06	1	1202	13.28	1
	2	1210	8.38	1	1212	9.18	1
	3	1215	6.78	1	1218	7.41	1
	4	1219	5.83	1	1222	6.38	1
	5	1222	5.20	1	1225	5.69	1
	6	1224	4.74	1	1228	5.20	1
	7	1227	4.40	1	1230	4.83	1
	8	1228	4.14	1	1232	4.54	1
	9	1230	3.94	1	1234	4.31	1
	10	1232	3.79	1	1236	4.12	2
	11	1233	3.67	1	1238	3.97	2
	12	1235	3.58	2	1239	3.85	2
	13	1236	3.51	2	1241	3.76	2
	14	1238	3.46	2	1242	3.68	3
	15	1239	3.43	2	1244	3.63	3
	16	1240	3.40	2	1245	3.59	3
	17	1242	3.39	3	1246	3.57	3
	18	1243	3.40	3	1248	3.56	3
	19	1244	3.41	3	1249	3.57	3
	20	1245	3.44	3	1250	3.59	3
	21	1247	3.48	3	1252	3.63	3
	22	1248	3.54	3	1253	3.69	3
	23	1250	3.61	3	1254	3.77	4
	24	1251	3.71	3	1256	3.87	4
	25	1253	3.83	3	1257	4.00	4
	26	1254	3.97	4	1259	4.17	4
	27	1256	4.15	4	1261	4.38	4
	28	1258	4.38	4	1263	4.65	4
	29	1260	4.68	4	1265	5.00	4
	30	1262	5.07	4	1268	5.48	4
	31	1265	5.62	4	1271	6.15	4
	32	1269	6.46	4	1275	7.16	4
	33	1274	7.92	4	1281	8.91	4
	34	1282	11.31	4	1290	12.98	4
	35	1290	20.84	4	1290	16.32	4

Table K-9. Raw to Scaled Score Look-up Table—Mathematics Grade 4—2022 to 2023

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
A	0	1200	30.19	1	1200	41.99	1
	1	1203	10.67	1	1200	17.60	1
	2	1211	7.35	1	1205	9.29	1
	3	1216	5.95	1	1211	6.95	1
	4	1219	5.16	1	1215	5.79	1
	5	1222	4.63	1	1219	5.08	1
	6	1224	4.26	1	1221	4.61	1
	7	1226	3.99	1	1223	4.27	1
	8	1227	3.77	1	1225	4.01	1
	9	1229	3.60	1	1227	3.81	1
	10	1230	3.47	1	1228	3.65	1
	11	1232	3.36	2	1230	3.52	1
	12	1233	3.27	2	1231	3.42	1
	13	1234	3.19	2	1232	3.34	2
	14	1235	3.14	2	1234	3.27	2
	15	1237	3.09	2	1235	3.22	2
	16	1238	3.07	2	1236	3.18	2
	17	1239	3.05	3	1237	3.16	2
	18	1240	3.05	3	1238	3.15	2
	19	1241	3.06	3	1240	3.16	3
	20	1242	3.09	3	1241	3.19	3
	21	1243	3.14	3	1242	3.23	3
	22	1245	3.21	3	1243	3.30	3
	23	1246	3.29	3	1245	3.38	3
	24	1247	3.41	3	1246	3.50	3
	25	1249	3.54	3	1247	3.65	3
	26	1250	3.72	3	1249	3.84	3
	27	1252	3.93	4	1251	4.08	4
	28	1254	4.20	4	1253	4.39	4
	29	1256	4.54	4	1255	4.79	4
	30	1258	4.99	4	1258	5.33	4
	31	1261	5.60	4	1261	6.11	4
	32	1265	6.50	4	1265	7.34	4
	33	1270	8.02	4	1272	9.71	4
	34	1278	11.50	4	1286	17.65	4
	35	1290	21.19	4	1290	23.03	4
B	0	1200	27.72	1	1200	47.12	1
	1	1202	10.18	1	1200	17.62	1
	2	1209	7.04	1	1207	9.41	1
	3	1214	5.71	1	1213	7.06	1
	4	1217	4.95	1	1217	5.88	1
	5	1219	4.46	1	1220	5.15	1
	6	1221	4.11	1	1223	4.66	1
	7	1223	3.85	1	1225	4.31	1
	8	1225	3.65	1	1227	4.04	1
	9	1226	3.49	1	1229	3.83	1
	10	1228	3.37	1	1230	3.67	1
	11	1229	3.27	1	1232	3.54	2
	12	1230	3.20	1	1233	3.43	2
	13	1231	3.14	1	1234	3.35	2
	14	1233	3.09	2	1236	3.29	2
	15	1234	3.06	2	1237	3.24	2
	16	1235	3.04	2	1238	3.22	2
	17	1236	3.04	2	1239	3.20	3
	18	1237	3.04	2	1241	3.21	3
	19	1238	3.06	2	1242	3.23	3

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
B	20	1239	3.08	3	1243	3.28	3
	21	1240	3.13	3	1244	3.34	3
	22	1242	3.18	3	1246	3.43	3
	23	1243	3.26	3	1247	3.55	3
	24	1244	3.35	3	1248	3.70	3
	25	1245	3.47	3	1250	3.88	3
	26	1247	3.62	3	1252	4.10	4
	27	1249	3.81	3	1254	4.39	4
	28	1250	4.05	3	1256	4.74	4
	29	1252	4.37	4	1258	5.20	4
	30	1254	4.79	4	1261	5.81	4
	31	1257	5.38	4	1265	6.67	4
	32	1261	6.27	4	1269	8.02	4
	33	1266	7.82	4	1277	10.59	4
	34	1274	11.39	4	1290	19.16	4
	35	1290	25.51	4	1290	19.25	4
C	0	1200	35.42	1	1200	34.57	1
	1	1207	10.77	1	1206	11.56	1
	2	1215	7.44	1	1214	7.92	1
	3	1219	6.05	1	1219	6.40	1
	4	1223	5.27	1	1223	5.53	1
	5	1225	4.76	1	1226	4.96	1
	6	1228	4.39	1	1228	4.55	1
	7	1230	4.11	1	1230	4.25	1
	8	1232	3.89	2	1232	4.01	2
	9	1233	3.71	2	1234	3.81	2
	10	1235	3.56	2	1235	3.65	2
	11	1236	3.43	2	1237	3.51	2
	12	1238	3.32	2	1238	3.39	2
	13	1239	3.23	3	1240	3.29	3
	14	1240	3.16	3	1241	3.21	3
	15	1241	3.09	3	1242	3.14	3
	16	1243	3.05	3	1243	3.09	3
	17	1244	3.01	3	1244	3.06	3
	18	1245	2.99	3	1246	3.04	3
	19	1246	2.99	3	1247	3.04	3
	20	1247	3.00	3	1248	3.05	3
	21	1248	3.04	3	1249	3.09	3
	22	1250	3.09	3	1250	3.15	3
	23	1251	3.18	4	1252	3.24	4
	24	1252	3.28	4	1253	3.35	4
	25	1254	3.42	4	1254	3.50	4
	26	1255	3.59	4	1256	3.68	4
	27	1257	3.81	4	1258	3.90	4
	28	1259	4.08	4	1259	4.17	4
	29	1261	4.41	4	1261	4.51	4
	30	1263	4.85	4	1264	4.96	4
	31	1266	5.44	4	1267	5.57	4
	32	1269	6.32	4	1270	6.48	4
	33	1274	7.83	4	1275	8.04	4
	34	1283	11.33	4	1284	11.68	4
	35	1290	17.36	4	1290	16.73	4

Table K-10. Raw to Scaled Score Look-up Table—Mathematics Grade 5—2022 to 2023

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
A	0	1200	24.06	1	1200	22.94	1
	1	1200	11.56	1	1200	12.89	1
	2	1207	7.76	1	1205	8.60	1
	3	1212	6.11	1	1210	6.73	1
	4	1216	5.16	1	1214	5.65	1
	5	1218	4.54	1	1217	4.93	1
	6	1221	4.10	1	1219	4.42	1
	7	1223	3.79	1	1221	4.05	1
	8	1224	3.55	1	1223	3.77	1
	9	1226	3.37	1	1225	3.56	1
	10	1227	3.24	1	1226	3.40	1
	11	1228	3.14	1	1228	3.28	1
	12	1230	3.07	1	1229	3.19	1
	13	1231	3.02	1	1230	3.14	1
	14	1232	3.00	2	1231	3.10	1
	15	1233	3.00	2	1233	3.10	2
	16	1235	3.02	2	1234	3.11	2
	17	1236	3.05	2	1235	3.15	2
	18	1237	3.10	2	1236	3.20	2
	19	1238	3.17	2	1238	3.28	2
	20	1239	3.25	2	1239	3.39	2
	21	1241	3.35	3	1240	3.52	3
	22	1242	3.47	3	1242	3.69	3
	23	1244	3.60	3	1244	3.88	3
	24	1245	3.74	3	1245	4.11	3
	25	1247	3.90	3	1247	4.38	3
	26	1249	4.09	3	1249	4.69	3
	27	1251	4.31	3	1252	5.04	3
	28	1253	4.57	4	1254	5.43	4
	29	1256	4.91	4	1257	5.89	4
	30	1258	5.35	4	1260	6.43	4
	31	1262	5.96	4	1264	7.13	4
	32	1266	6.86	4	1269	8.12	4
	33	1271	8.38	4	1275	9.80	4
	34	1280	11.91	4	1286	13.81	4
	35	1290	21.01	4	1290	18.44	4
B	0	1200	27.87	1	1200	26.42	1
	1	1202	12.02	1	1200	14.06	1
	2	1211	8.32	1	1207	9.49	1
	3	1216	6.75	1	1213	7.58	1
	4	1220	5.86	1	1217	6.50	1
	5	1223	5.28	1	1221	5.80	1
	6	1226	4.89	1	1224	5.32	1
	7	1228	4.60	1	1226	4.97	1
	8	1230	4.39	1	1228	4.71	1
	9	1232	4.22	2	1230	4.51	1
	10	1233	4.09	2	1232	4.36	2
	11	1235	3.99	2	1234	4.24	2
	12	1237	3.91	2	1235	4.14	2
	13	1238	3.84	2	1237	4.06	2
	14	1239	3.79	2	1239	4.00	2
	15	1241	3.74	3	1240	3.95	3
	16	1243	3.72	3	1242	3.92	3
	17	1244	3.70	3	1243	3.90	3
	18	1245	3.69	3	1245	3.90	3
	19	1247	3.70	3	1246	3.91	3

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
B	20	1248	3.72	3	1248	3.94	3
	21	1250	3.76	3	1249	3.98	3
	22	1251	3.81	3	1251	4.05	3
	23	1253	3.88	4	1252	4.14	3
	24	1254	3.97	4	1254	4.25	4
	25	1256	4.08	4	1256	4.39	4
	26	1257	4.23	4	1258	4.57	4
	27	1259	4.40	4	1260	4.79	4
	28	1261	4.63	4	1262	5.08	4
	29	1264	4.93	4	1264	5.45	4
	30	1266	5.33	4	1267	5.94	4
	31	1269	5.90	4	1271	6.64	4
	32	1273	6.77	4	1275	7.70	4
	33	1278	8.30	4	1281	9.57	4
	34	1287	11.95	4	1290	14.04	4
	35	1290	15.94	4	1290	15.39	4
C	0	1200	30.65	1	1200	30.34	1
	1	1205	11.79	1	1203	12.55	1
	2	1213	8.18	1	1212	8.63	1
	3	1219	6.65	1	1218	6.97	1
	4	1222	5.77	1	1222	6.04	1
	5	1225	5.20	1	1225	5.44	1
	6	1228	4.80	1	1228	5.02	1
	7	1230	4.51	1	1230	4.71	1
	8	1232	4.28	2	1232	4.47	2
	9	1234	4.09	2	1234	4.28	2
	10	1236	3.94	2	1236	4.13	2
	11	1237	3.81	2	1238	4.00	2
	12	1239	3.71	2	1239	3.89	2
	13	1240	3.62	3	1241	3.81	3
	14	1242	3.56	3	1242	3.74	3
	15	1243	3.50	3	1244	3.69	3
	16	1244	3.47	3	1245	3.66	3
	17	1246	3.45	3	1246	3.65	3
	18	1247	3.44	3	1248	3.66	3
	19	1248	3.46	3	1249	3.68	3
	20	1250	3.48	3	1251	3.72	3
	21	1251	3.52	3	1252	3.78	3
	22	1252	3.59	3	1254	3.87	4
	23	1254	3.67	4	1255	3.97	4
	24	1255	3.77	4	1257	4.09	4
	25	1257	3.90	4	1259	4.24	4
	26	1258	4.06	4	1260	4.43	4
	27	1260	4.26	4	1262	4.66	4
	28	1262	4.51	4	1265	4.95	4
	29	1264	4.84	4	1267	5.31	4
	30	1267	5.27	4	1270	5.79	4
	31	1270	5.88	4	1273	6.45	4
	32	1274	6.80	4	1277	7.45	4
	33	1279	8.39	4	1283	9.17	4
	34	1288	12.08	4	1290	13.16	4
	35	1290	15.39	4	1290	13.69	4

Table K-11. Raw to Scaled Score Look-up Table—Mathematics Grade 6—2022 to 2023

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
A	0	1200	33.52	1	1200	31.40	1
	1	1209	8.85	1	1208	8.27	1
	2	1215	6.14	1	1214	5.70	1
	3	1219	5.00	1	1217	4.64	1
	4	1222	4.35	1	1220	4.04	1
	5	1224	3.92	1	1222	3.65	1
	6	1226	3.61	1	1223	3.38	1
	7	1227	3.39	1	1225	3.18	1
	8	1229	3.21	1	1226	3.03	1
	9	1230	3.07	1	1228	2.92	1
	10	1231	2.96	1	1229	2.83	1
	11	1233	2.88	2	1230	2.77	1
	12	1234	2.81	2	1231	2.72	1
	13	1235	2.76	2	1232	2.69	1
	14	1236	2.73	2	1233	2.67	2
	15	1237	2.71	2	1234	2.66	2
	16	1238	2.70	2	1235	2.67	2
	17	1239	2.70	3	1236	2.68	2
	18	1240	2.72	3	1237	2.70	2
	19	1241	2.74	3	1238	2.74	2
	20	1242	2.78	3	1239	2.78	3
	21	1243	2.84	3	1240	2.84	3
	22	1244	2.90	3	1241	2.91	3
	23	1245	2.98	3	1242	3.00	3
	24	1246	3.08	3	1243	3.11	3
	25	1247	3.20	3	1245	3.23	3
	26	1249	3.34	3	1246	3.39	3
	27	1250	3.52	3	1248	3.58	3
	28	1252	3.75	4	1249	3.81	3
	29	1254	4.04	4	1251	4.11	4
	30	1256	4.42	4	1253	4.51	4
	31	1258	4.96	4	1256	5.05	4
	32	1261	5.76	4	1259	5.87	4
	33	1266	7.16	4	1264	7.25	4
	34	1274	10.41	4	1271	10.44	4
	35	1290	25.84	4	1290	28.42	4
B	0	1200	30.30	1	1200	33.92	1
	1	1207	8.03	1	1206	10.03	1
	2	1213	5.58	1	1214	6.65	1
	3	1217	4.55	1	1218	5.32	1
	4	1219	3.96	1	1221	4.58	1
	5	1221	3.58	1	1223	4.10	1
	6	1223	3.31	1	1225	3.76	1
	7	1224	3.11	1	1227	3.52	1
	8	1225	2.96	1	1228	3.33	1
	9	1227	2.84	1	1230	3.18	1
	10	1228	2.75	1	1231	3.06	1
	11	1229	2.68	1	1232	2.97	1
	12	1230	2.63	1	1233	2.90	2
	13	1231	2.59	1	1234	2.84	2
	14	1232	2.56	1	1235	2.81	2
	15	1233	2.55	2	1237	2.78	2
	16	1234	2.54	2	1238	2.78	2
	17	1235	2.55	2	1239	2.78	3
	18	1236	2.57	2	1240	2.80	3
	19	1237	2.60	2	1241	2.83	3

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
B	20	1237	2.63	2	1242	2.88	3
	21	1238	2.68	2	1243	2.94	3
	22	1240	2.75	3	1244	3.01	3
	23	1241	2.82	3	1245	3.10	3
	24	1242	2.92	3	1246	3.21	3
	25	1243	3.03	3	1248	3.34	3
	26	1244	3.17	3	1249	3.50	3
	27	1246	3.34	3	1251	3.70	4
	28	1247	3.56	3	1252	3.94	4
	29	1249	3.84	3	1254	4.24	4
	30	1251	4.21	4	1256	4.65	4
	31	1253	4.72	4	1259	5.21	4
	32	1257	5.50	4	1262	6.05	4
	33	1261	6.86	4	1267	7.50	4
	34	1268	10.05	4	1275	10.93	4
	35	1290	32.95	4	1290	24.76	4
C	0	1200	38.69	1	1200	36.40	1
	1	1211	9.01	1	1210	9.39	1
	2	1218	6.28	1	1217	6.48	1
	3	1222	5.13	1	1221	5.26	1
	4	1225	4.48	1	1224	4.57	1
	5	1227	4.05	1	1226	4.12	1
	6	1229	3.75	1	1228	3.80	1
	7	1231	3.52	1	1230	3.56	1
	8	1232	3.35	1	1232	3.38	1
	9	1234	3.22	2	1233	3.24	2
	10	1235	3.11	2	1234	3.13	2
	11	1236	3.03	2	1236	3.05	2
	12	1237	2.96	2	1237	2.98	2
	13	1238	2.91	2	1238	2.93	2
	14	1239	2.88	3	1239	2.89	3
	15	1240	2.85	3	1240	2.86	3
	16	1242	2.84	3	1241	2.85	3
	17	1243	2.84	3	1242	2.85	3
	18	1244	2.85	3	1243	2.85	3
	19	1245	2.87	3	1244	2.87	3
	20	1246	2.90	3	1245	2.90	3
	21	1247	2.94	3	1246	2.94	3
	22	1248	3.00	3	1248	3.00	3
	23	1249	3.07	3	1249	3.07	3
	24	1250	3.16	3	1250	3.16	3
	25	1252	3.27	4	1251	3.27	4
	26	1253	3.41	4	1253	3.41	4
	27	1255	3.58	4	1254	3.58	4
	28	1256	3.80	4	1256	3.80	4
	29	1258	4.09	4	1258	4.09	4
	30	1260	4.47	4	1260	4.47	4
	31	1263	5.00	4	1262	5.01	4
	32	1266	5.79	4	1266	5.82	4
	33	1271	7.17	4	1270	7.21	4
	34	1278	10.35	4	1278	10.43	4
	35	1290	20.72	4	1290	21.22	4

Table K-12. Raw to Scaled Score Look-up Table—Mathematics Grade 7—2022 to 2023

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
A	0	1200	19.84	1	1200	19.73	1
	1	1200	10.98	1	1200	12.81	1
	2	1206	7.52	1	1203	8.59	1
	3	1211	6.04	1	1209	6.82	1
	4	1214	5.19	1	1213	5.81	1
	5	1217	4.64	1	1216	5.16	1
	6	1219	4.25	1	1218	4.70	1
	7	1221	3.96	1	1220	4.36	1
	8	1223	3.74	1	1222	4.11	1
	9	1224	3.58	1	1224	3.91	1
	10	1226	3.45	1	1225	3.76	1
	11	1227	3.35	1	1227	3.64	1
	12	1228	3.27	1	1228	3.55	1
	13	1230	3.21	1	1230	3.48	1
	14	1231	3.18	1	1231	3.44	1
	15	1232	3.16	1	1232	3.41	1
	16	1233	3.15	1	1234	3.41	2
	17	1234	3.16	2	1235	3.42	2
	18	1236	3.19	2	1236	3.46	2
	19	1237	3.24	2	1238	3.52	2
	20	1238	3.30	2	1239	3.60	2
	21	1239	3.38	2	1241	3.70	3
	22	1241	3.48	3	1242	3.83	3
	23	1242	3.59	3	1244	3.98	3
	24	1244	3.74	3	1245	4.16	3
	25	1245	3.90	3	1247	4.37	3
	26	1247	4.10	3	1249	4.62	3
	27	1249	4.34	3	1251	4.92	3
	28	1251	4.62	3	1254	5.28	4
	29	1254	4.98	4	1256	5.74	4
	30	1256	5.44	4	1259	6.34	4
	31	1259	6.07	4	1263	7.16	4
	32	1263	7.01	4	1268	8.38	4
	33	1268	8.59	4	1275	10.45	4
	34	1277	12.23	4	1286	15.23	4
	35	1290	25.06	4	1290	20.60	4
B	0	1200	21.08	1	1200	21.97	1
	1	1200	11.14	1	1200	13.59	1
	2	1208	7.85	1	1205	9.23	1
	3	1213	6.45	1	1211	7.43	1
	4	1216	5.65	1	1215	6.41	1
	5	1219	5.13	1	1218	5.75	1
	6	1221	4.76	1	1221	5.28	1
	7	1224	4.49	1	1224	4.94	1
	8	1226	4.28	1	1226	4.67	1
	9	1227	4.12	1	1228	4.46	1
	10	1229	3.99	1	1229	4.29	1
	11	1231	3.88	1	1231	4.16	1
	12	1232	3.80	1	1233	4.05	1
	13	1234	3.73	2	1234	3.96	2
	14	1235	3.68	2	1236	3.90	2
	15	1236	3.64	2	1237	3.86	2
	16	1238	3.62	2	1239	3.83	2
	17	1239	3.60	2	1240	3.82	3

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
B	18	1240	3.61	3	1242	3.83	3
	19	1242	3.62	3	1243	3.85	3
	20	1243	3.65	3	1244	3.90	3
	21	1244	3.69	3	1246	3.96	3
	22	1246	3.75	3	1247	4.04	3
	23	1247	3.83	3	1249	4.15	3
	24	1249	3.92	3	1251	4.28	3
	25	1250	4.04	3	1252	4.45	3
	26	1252	4.19	3	1254	4.65	4
	27	1254	4.38	4	1256	4.90	4
	28	1256	4.62	4	1259	5.22	4
	29	1258	4.92	4	1261	5.63	4
	30	1261	5.33	4	1264	6.19	4
	31	1264	5.89	4	1268	6.96	4
	32	1267	6.74	4	1272	8.13	4
	33	1272	8.20	4	1279	10.15	4
	34	1281	11.58	4	1289	14.88	4
	35	1290	21.33	4	1290	18.01	4
C	0	1200	25.93	1	1200	26.57	1
	1	1204	11.20	1	1202	12.93	1
	2	1212	7.87	1	1211	8.90	1
	3	1217	6.46	1	1217	7.19	1
	4	1221	5.65	1	1221	6.19	1
	5	1224	5.12	1	1224	5.52	1
	6	1227	4.72	1	1227	5.03	1
	7	1229	4.42	1	1229	4.64	1
	8	1231	4.16	1	1232	4.33	1
	9	1233	3.95	1	1233	4.07	1
	10	1235	3.77	2	1235	3.86	2
	11	1236	3.61	2	1237	3.69	2
	12	1238	3.48	2	1238	3.54	2
	13	1239	3.38	2	1239	3.43	2
	14	1240	3.29	3	1241	3.33	3
	15	1242	3.22	3	1242	3.27	3
	16	1243	3.17	3	1243	3.22	3
	17	1244	3.14	3	1245	3.19	3
	18	1245	3.12	3	1246	3.18	3
	19	1246	3.12	3	1247	3.19	3
	20	1248	3.14	3	1248	3.21	3
	21	1249	3.18	3	1250	3.26	3
	22	1250	3.23	3	1251	3.33	3
	23	1251	3.30	3	1252	3.41	3
	24	1253	3.40	3	1254	3.53	4
	25	1254	3.52	4	1255	3.67	4
	26	1256	3.68	4	1257	3.85	4
	27	1257	3.87	4	1258	4.07	4
	28	1259	4.11	4	1260	4.35	4
	29	1261	4.42	4	1262	4.71	4
	30	1264	4.84	4	1265	5.20	4
	31	1266	5.42	4	1268	5.89	4
	32	1270	6.31	4	1272	6.94	4
	33	1275	7.88	4	1278	8.80	4
	34	1284	11.58	4	1287	13.25	4
	35	1290	18.94	4	1290	17.92	4

Table K-13. Raw to Scaled Score Look-up Table—Mathematics Grade 8—2022 to 2023

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
A	0	1200	25.07	1	1200	26.12	1
	1	1201	11.18	1	1200	12.51	1
	2	1209	7.45	1	1209	7.99	1
	3	1214	5.85	1	1214	6.11	1
	4	1217	4.94	1	1217	5.08	1
	5	1220	4.36	1	1220	4.44	1
	6	1222	3.97	1	1222	4.02	1
	7	1224	3.69	1	1224	3.72	1
	8	1226	3.50	1	1226	3.52	1
	9	1227	3.36	1	1227	3.37	1
	10	1228	3.26	1	1228	3.26	1
	11	1230	3.19	1	1230	3.19	1
	12	1231	3.14	1	1231	3.14	1
	13	1232	3.11	1	1232	3.11	1
	14	1233	3.10	1	1233	3.10	1
	15	1235	3.10	2	1235	3.10	2
	16	1236	3.12	2	1236	3.11	2
	17	1237	3.14	2	1237	3.14	2
	18	1238	3.18	2	1238	3.18	2
	19	1239	3.23	2	1240	3.23	3
	20	1241	3.29	3	1241	3.30	3
	21	1242	3.36	3	1242	3.37	3
	22	1243	3.44	3	1243	3.47	3
	23	1245	3.54	3	1245	3.58	3
	24	1246	3.66	3	1246	3.71	3
	25	1247	3.80	3	1248	3.87	3
	26	1249	3.97	3	1250	4.06	3
	27	1251	4.18	4	1251	4.29	4
	28	1253	4.43	4	1253	4.58	4
	29	1255	4.75	4	1256	4.95	4
	30	1257	5.18	4	1258	5.44	4
	31	1260	5.76	4	1262	6.13	4
	32	1264	6.64	4	1266	7.16	4
	33	1269	8.17	4	1271	8.95	4
	34	1278	11.74	4	1281	13.20	4
	35	1290	24.06	4	1290	22.19	4
B	0	1200	30.01	1	1200	25.56	1
	1	1206	10.83	1	1202	11.78	1
	2	1214	7.62	1	1210	8.20	1
	3	1219	6.26	1	1215	6.66	1
	4	1222	5.47	1	1219	5.75	1
	5	1225	4.95	1	1222	5.14	1
	6	1227	4.57	1	1224	4.69	1
	7	1229	4.29	1	1227	4.35	1
	8	1231	4.07	1	1229	4.08	1
	9	1233	3.90	1	1230	3.86	1
	10	1235	3.76	2	1232	3.68	1
	11	1236	3.65	2	1233	3.53	1
	12	1237	3.56	2	1235	3.41	2
	13	1239	3.49	2	1236	3.31	2
	14	1240	3.44	3	1237	3.24	2
	15	1241	3.40	3	1238	3.18	2
	16	1242	3.37	3	1240	3.13	3
	17	1244	3.35	3	1241	3.11	3
	18	1245	3.35	3	1242	3.10	3
	19	1246	3.36	3	1243	3.11	3
	20	1247	3.39	3	1244	3.14	3

continued

Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
B	21	1248	3.42	3	1245	3.18	3
	22	1250	3.48	3	1247	3.25	3
	23	1251	3.54	4	1248	3.34	3
	24	1252	3.63	4	1249	3.46	3
	25	1254	3.75	4	1251	3.61	4
	26	1255	3.89	4	1252	3.80	4
	27	1257	4.06	4	1254	4.03	4
	28	1259	4.29	4	1256	4.32	4
	29	1261	4.58	4	1258	4.70	4
	30	1263	4.97	4	1260	5.19	4
	31	1266	5.51	4	1263	5.87	4
	32	1269	6.34	4	1267	6.88	4
	33	1274	7.75	4	1273	8.59	4
	34	1282	11.05	4	1282	12.52	4
	35	1290	19.39	4	1290	20.66	4
C	0	1200	27.37	1	1200	29.75	1
	1	1204	10.95	1	1206	11.28	1
	2	1212	7.67	1	1214	7.89	1
	3	1217	6.24	1	1219	6.45	1
	4	1220	5.41	1	1222	5.63	1
	5	1223	4.85	1	1225	5.08	1
	6	1226	4.44	1	1227	4.69	1
	7	1228	4.13	1	1229	4.40	1
	8	1229	3.88	1	1231	4.18	1
	9	1231	3.68	1	1233	4.00	1
	10	1232	3.52	1	1235	3.86	2
	11	1233	3.39	1	1236	3.74	2
	12	1235	3.28	2	1238	3.65	2
	13	1236	3.19	2	1239	3.58	2
	14	1237	3.12	2	1240	3.52	3
	15	1239	3.07	2	1241	3.48	3
	16	1240	3.03	3	1243	3.46	3
	17	1241	3.01	3	1244	3.44	3
	18	1242	3.00	3	1245	3.44	3
	19	1243	3.01	3	1246	3.46	3
	20	1244	3.04	3	1248	3.48	3
	21	1245	3.08	3	1249	3.53	3
	22	1246	3.14	3	1250	3.58	3
	23	1248	3.22	3	1252	3.66	4
	24	1249	3.32	3	1253	3.76	4
	25	1250	3.45	3	1255	3.88	4
	26	1252	3.62	4	1256	4.03	4
	27	1253	3.82	4	1258	4.23	4
	28	1255	4.08	4	1260	4.47	4
	29	1257	4.41	4	1262	4.79	4
	30	1259	4.85	4	1264	5.21	4
	31	1262	5.45	4	1267	5.80	4
	32	1266	6.36	4	1271	6.70	4
	33	1271	7.91	4	1276	8.24	4
	34	1279	11.56	4	1285	11.82	4
	35	1290	22.74	4	1290	17.95	4

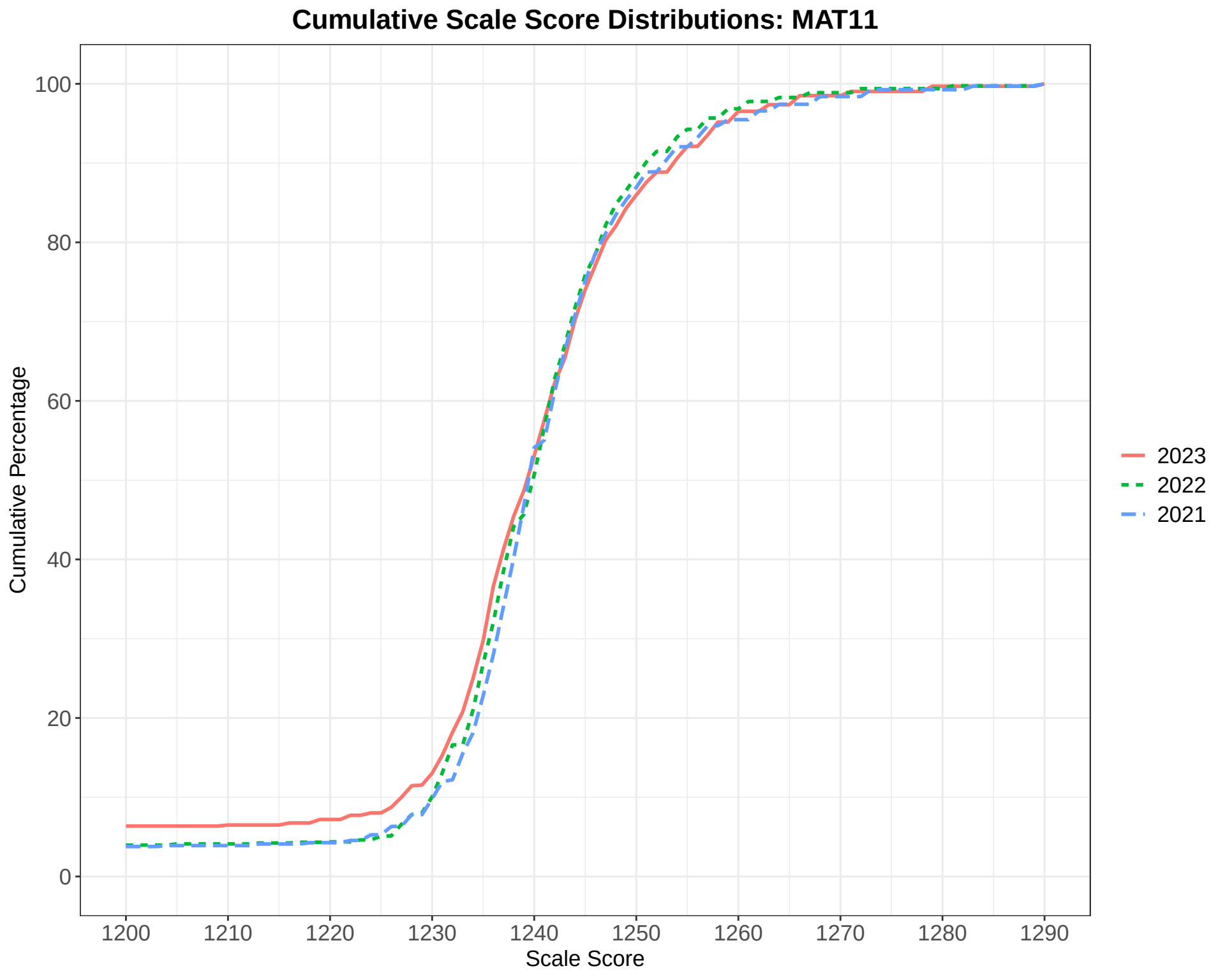
Table K-14. Raw to Scaled Score Look-up Table—Mathematics Grade HS—2022 to 2023

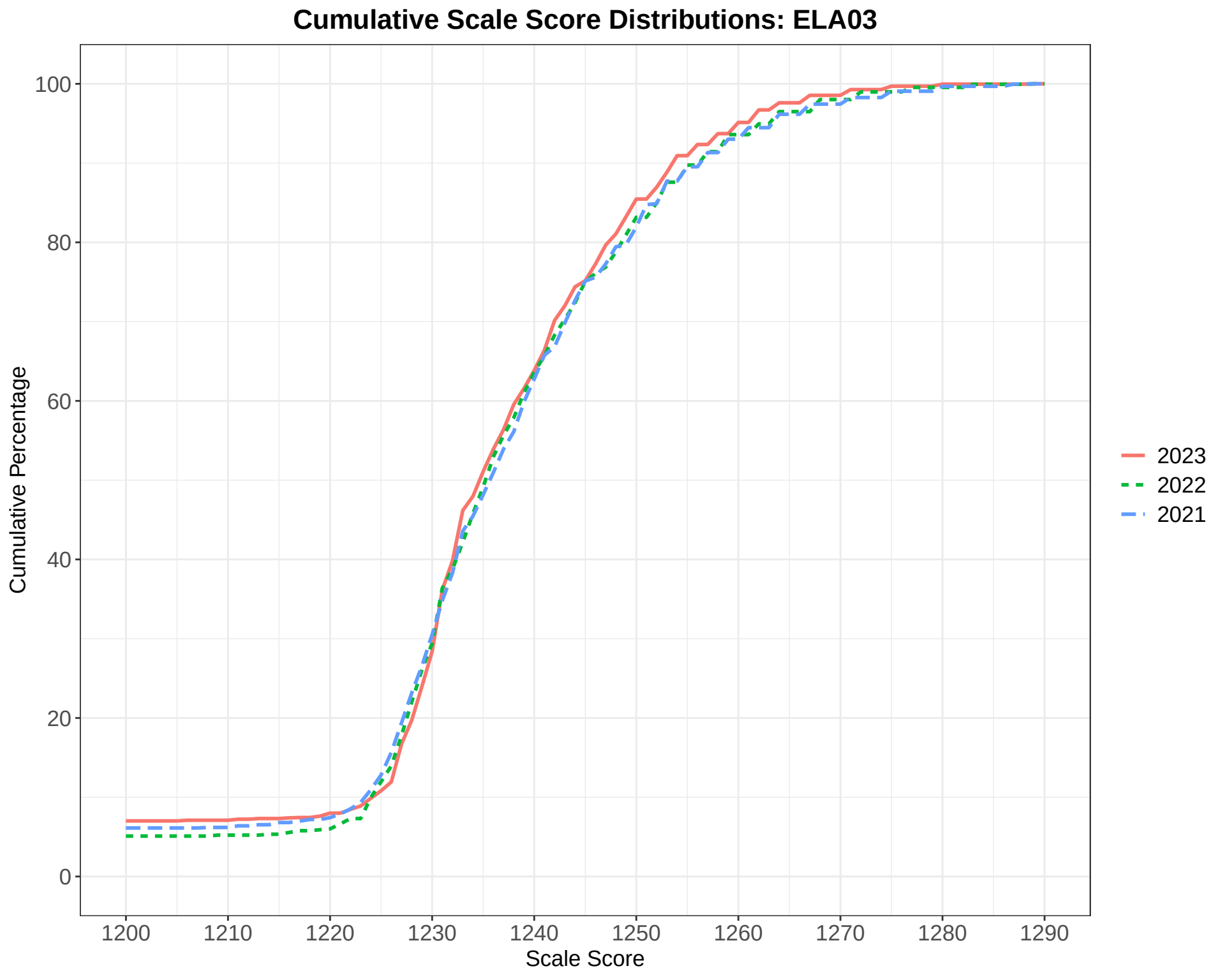
Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
A	0	1200	33.41	1	1200	32.42	1
	1	1210	8.18	1	1209	8.71	1
	2	1216	5.68	1	1215	5.95	1
	3	1219	4.60	1	1219	4.79	1
	4	1222	3.99	1	1221	4.12	1
	5	1224	3.58	1	1224	3.68	1
	6	1226	3.29	1	1225	3.38	1
	7	1227	3.08	1	1227	3.15	1
	8	1228	2.92	1	1228	2.98	1
	9	1230	2.80	1	1229	2.86	1
	10	1231	2.71	1	1231	2.76	1
	11	1232	2.64	1	1232	2.69	1
	12	1233	2.59	1	1233	2.63	1
	13	1234	2.55	1	1234	2.60	1
	14	1235	2.53	2	1235	2.58	2
	15	1236	2.52	2	1236	2.57	2
	16	1236	2.52	2	1237	2.57	2
	17	1237	2.53	2	1237	2.58	2
	18	1238	2.55	2	1238	2.61	2
	19	1239	2.58	2	1239	2.64	2
	20	1240	2.62	3	1240	2.69	3
	21	1241	2.67	3	1241	2.74	3
	22	1242	2.73	3	1242	2.81	3
	23	1243	2.80	3	1244	2.89	3
	24	1244	2.89	3	1245	2.99	3
	25	1246	3.00	3	1246	3.12	3
	26	1247	3.13	3	1247	3.26	3
	27	1248	3.29	3	1249	3.45	3
	28	1250	3.49	4	1250	3.68	4
	29	1251	3.74	4	1252	3.97	4
	30	1253	4.07	4	1254	4.37	4
	31	1256	4.53	4	1257	4.92	4
	32	1259	5.22	4	1260	5.76	4
	33	1263	6.39	4	1265	7.22	4
	34	1269	9.11	4	1273	10.68	4
	35	1290	35.79	4	1290	29.63	4
B	0	1200	39.36	1	1200	33.27	1
	1	1212	9.55	1	1206	11.23	1
	2	1219	6.60	1	1214	7.30	1
	3	1223	5.37	1	1219	5.74	1
	4	1226	4.67	1	1222	4.87	1
	5	1228	4.21	1	1224	4.31	1
	6	1231	3.89	1	1226	3.93	1
	7	1232	3.64	1	1228	3.64	1
	8	1234	3.45	1	1230	3.42	1
	9	1235	3.28	2	1231	3.24	1
	10	1237	3.14	2	1233	3.10	1
	11	1238	3.02	2	1234	2.99	1
	12	1239	2.91	2	1235	2.90	2
	13	1240	2.82	3	1236	2.82	2
	14	1241	2.73	3	1237	2.76	2
	15	1242	2.65	3	1238	2.72	2
	16	1243	2.58	3	1239	2.70	2

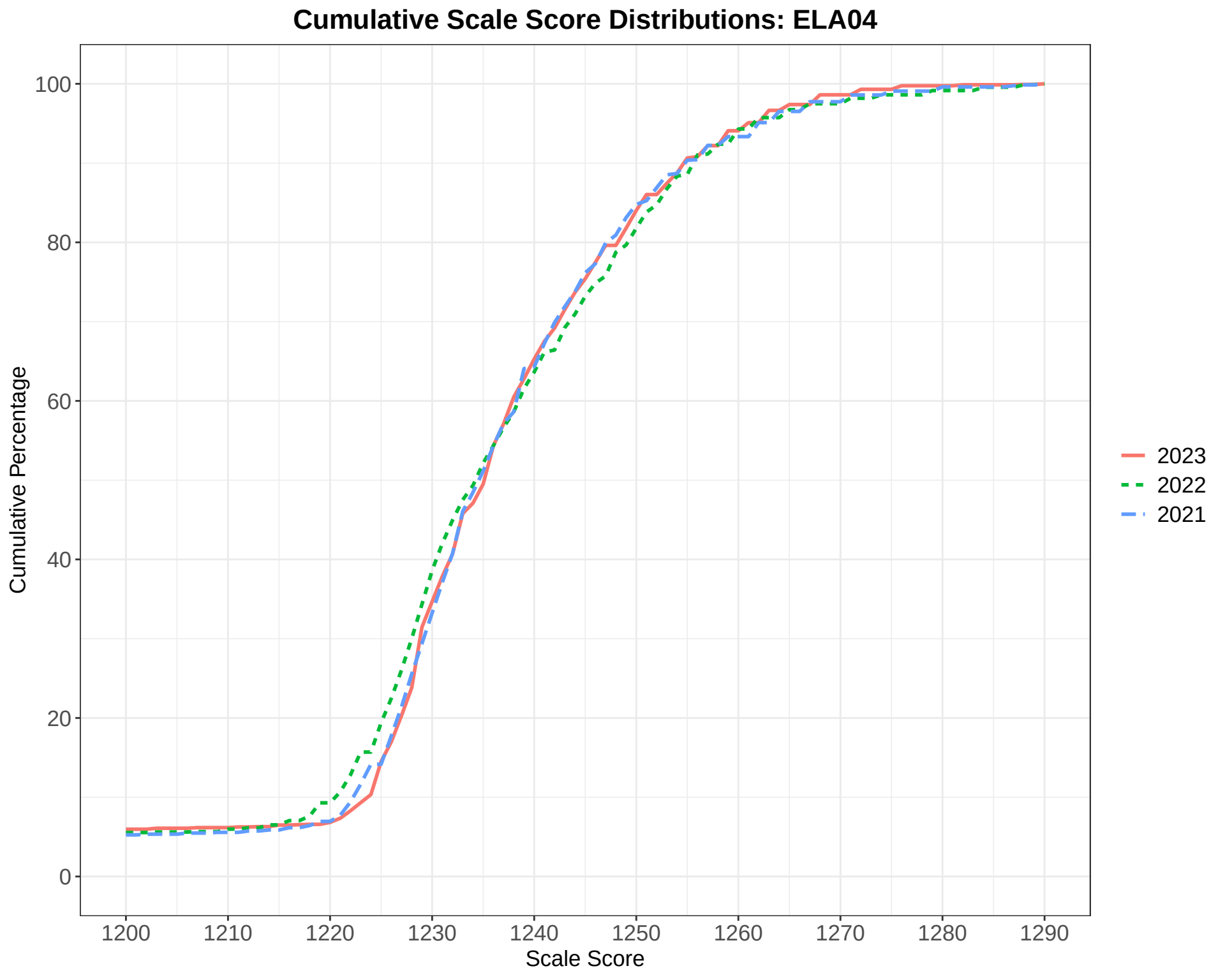
continued

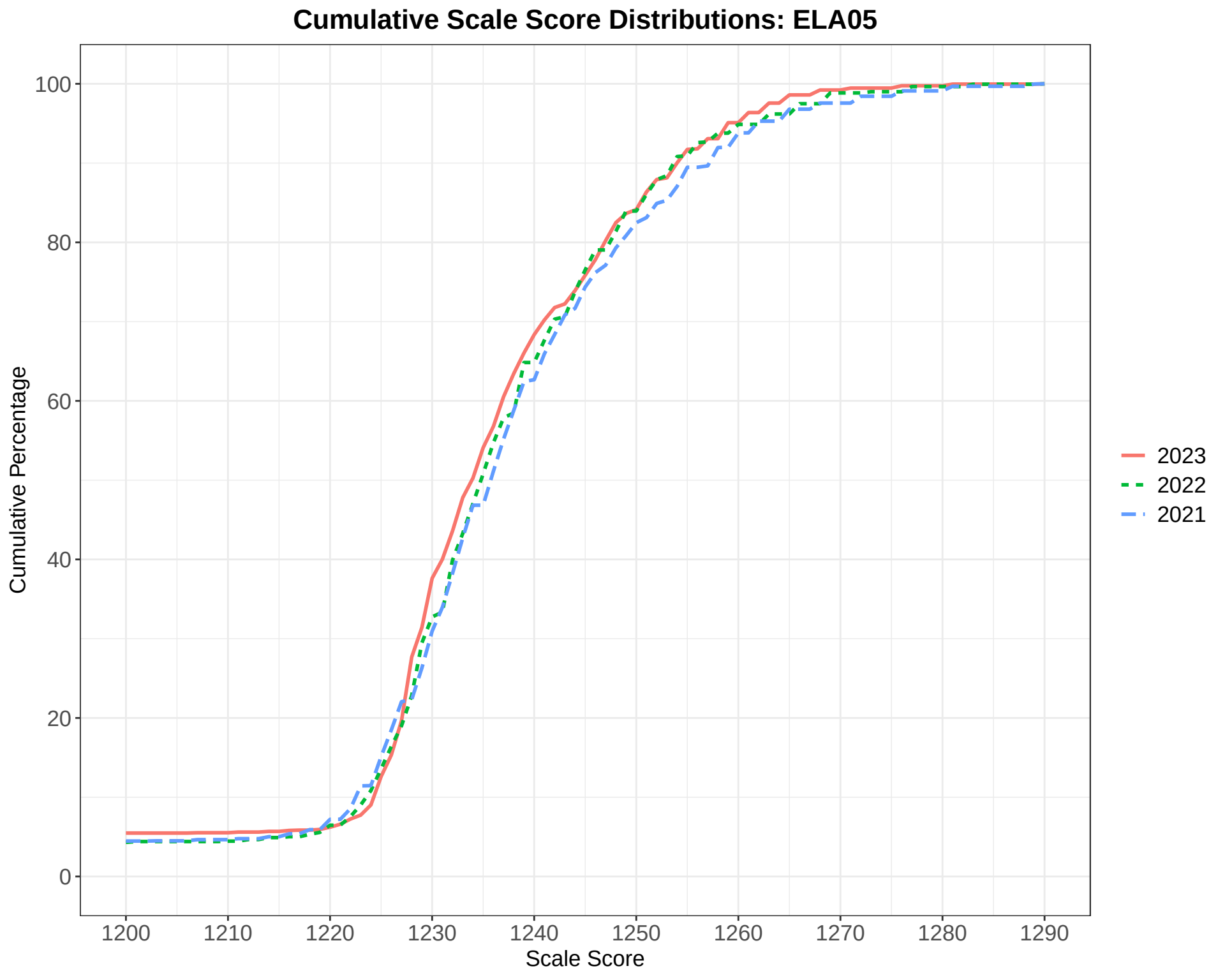
Path	Raw Score	2022			2023		
		Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
B	17	1244	2.53	3	1240	2.68	3
	18	1245	2.49	3	1241	2.69	3
	19	1246	2.47	3	1242	2.70	3
	20	1247	2.46	3	1243	2.74	3
	21	1248	2.48	3	1245	2.80	3
	22	1249	2.52	3	1246	2.87	3
	23	1250	2.59	4	1247	2.97	3
	24	1251	2.68	4	1248	3.10	3
	25	1252	2.81	4	1249	3.26	3
	26	1254	2.98	4	1251	3.46	4
	27	1255	3.20	4	1252	3.71	4
	28	1257	3.47	4	1254	4.03	4
	29	1258	3.82	4	1256	4.44	4
	30	1260	4.28	4	1259	4.99	4
	31	1263	4.90	4	1262	5.75	4
	32	1266	5.82	4	1266	6.92	4
	33	1271	7.38	4	1272	9.01	4
	34	1279	10.96	4	1283	14.20	4
	35	1290	22.49	4	1290	22.55	4
C	0	1200	33.66	1	1200	36.75	1
	1	1209	9.27	1	1208	11.33	1
	2	1216	6.39	1	1217	7.50	1
	3	1220	5.17	1	1221	5.96	1
	4	1223	4.47	1	1225	5.10	1
	5	1225	4.02	1	1227	4.55	1
	6	1227	3.69	1	1230	4.16	1
	7	1229	3.45	1	1232	3.87	1
	8	1230	3.26	1	1233	3.64	1
	9	1232	3.11	1	1235	3.45	2
	10	1233	2.98	1	1236	3.29	2
	11	1234	2.88	1	1238	3.15	2
	12	1235	2.80	2	1239	3.02	2
	13	1236	2.74	2	1240	2.91	3
	14	1237	2.69	2	1241	2.81	3
	15	1238	2.65	2	1242	2.72	3
	16	1239	2.63	2	1243	2.65	3
	17	1240	2.62	3	1244	2.59	3
	18	1241	2.62	3	1245	2.54	3
	19	1242	2.64	3	1246	2.52	3
	20	1243	2.67	3	1247	2.52	3
	21	1244	2.72	3	1248	2.55	3
	22	1245	2.79	3	1249	2.60	3
	23	1247	2.88	3	1251	2.68	4
	24	1248	2.99	3	1252	2.79	4
	25	1249	3.13	3	1253	2.95	4
	26	1250	3.30	4	1254	3.16	4
	27	1252	3.51	4	1256	3.43	4
	28	1254	3.78	4	1257	3.77	4
	29	1255	4.11	4	1259	4.22	4
	30	1258	4.55	4	1262	4.81	4
	31	1260	5.16	4	1265	5.64	4
	32	1264	6.06	4	1269	6.90	4
	33	1269	7.60	4	1275	9.11	4
	34	1277	11.20	4	1286	14.45	4
	35	1290	24.64	4	1290	20.13	4

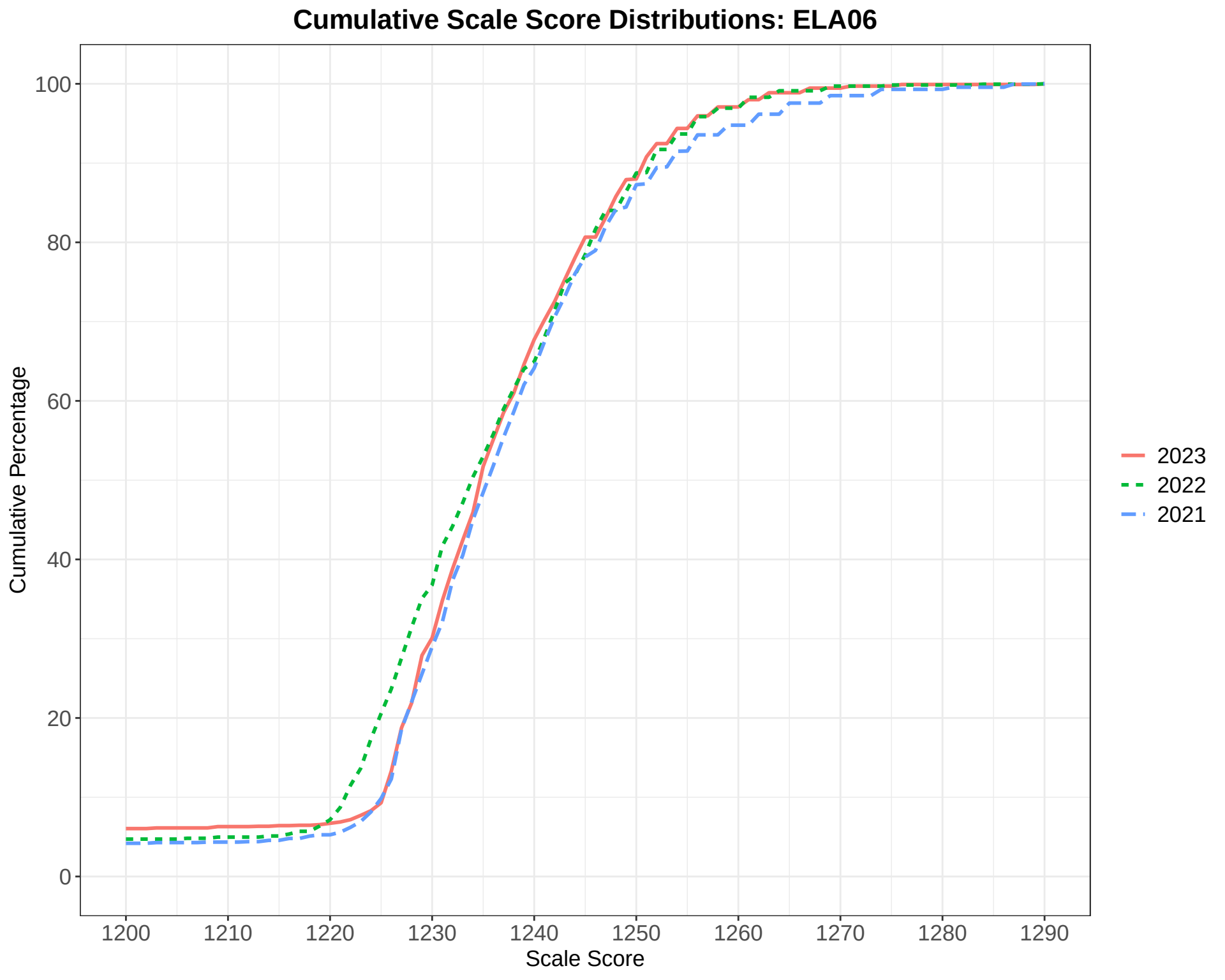
APPENDIX L
SCORE DISTRIBUTIONS

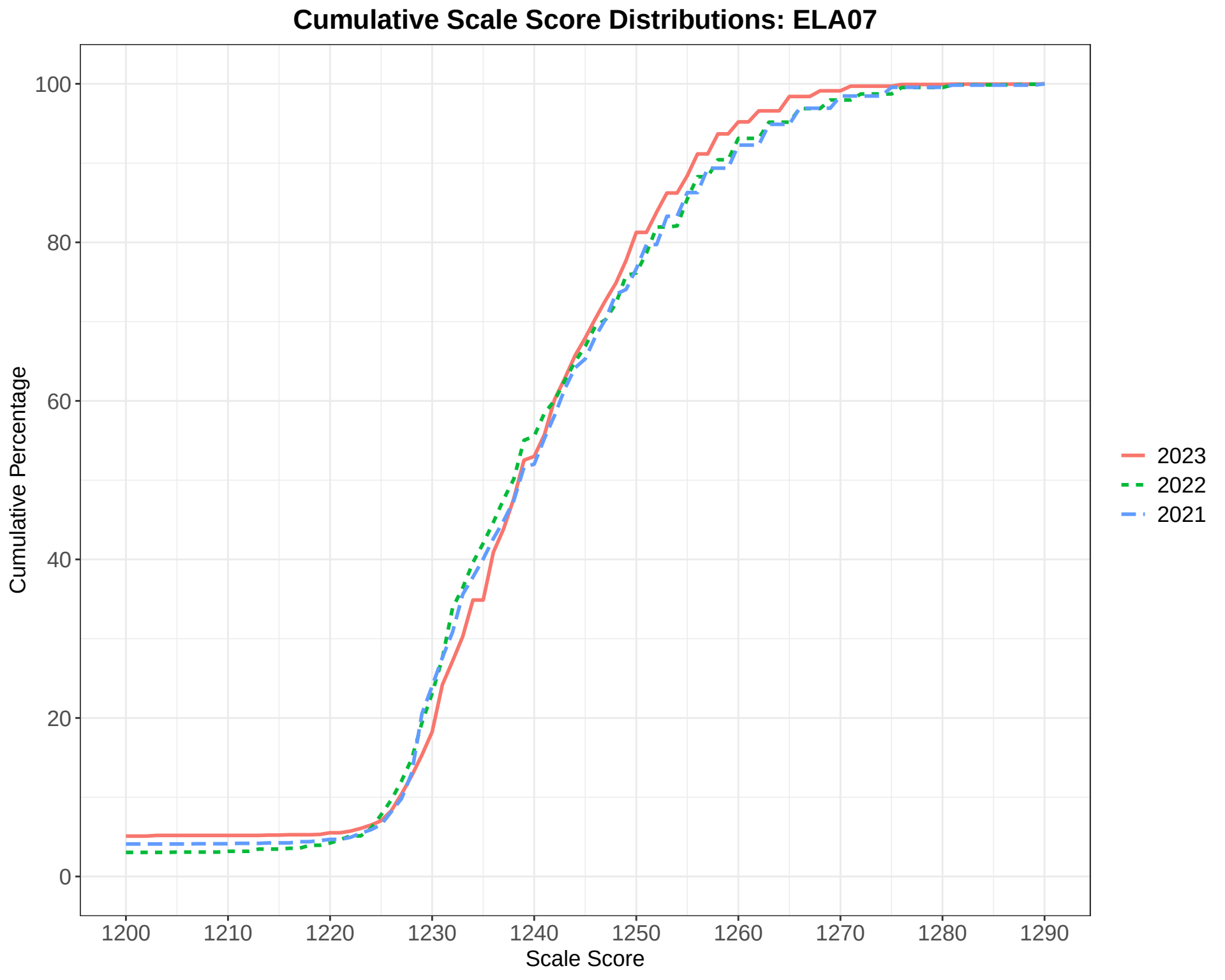

Cumulative Scale Score Distributions: MAT11
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021

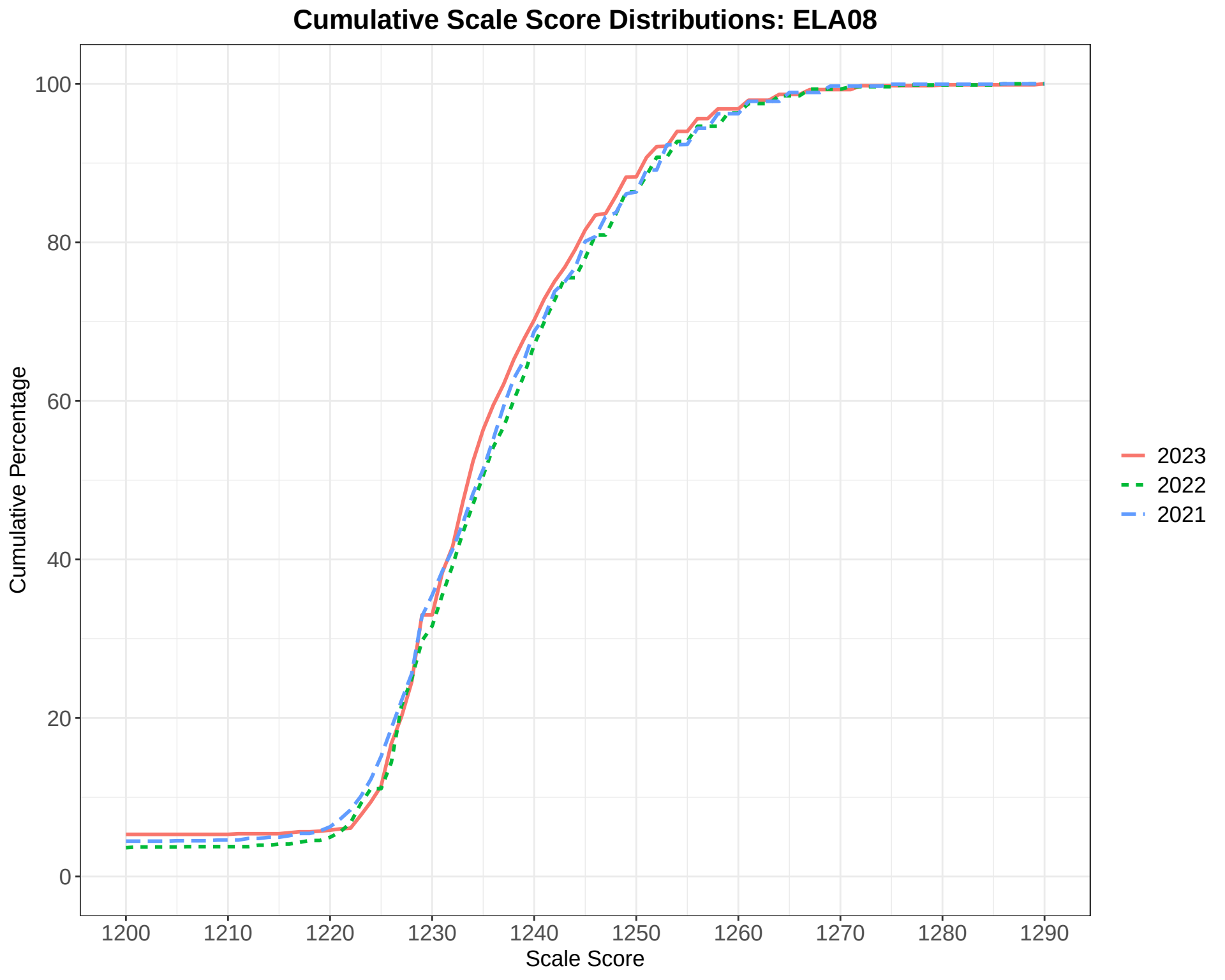

Cumulative Scale Score Distributions: ELA03
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021

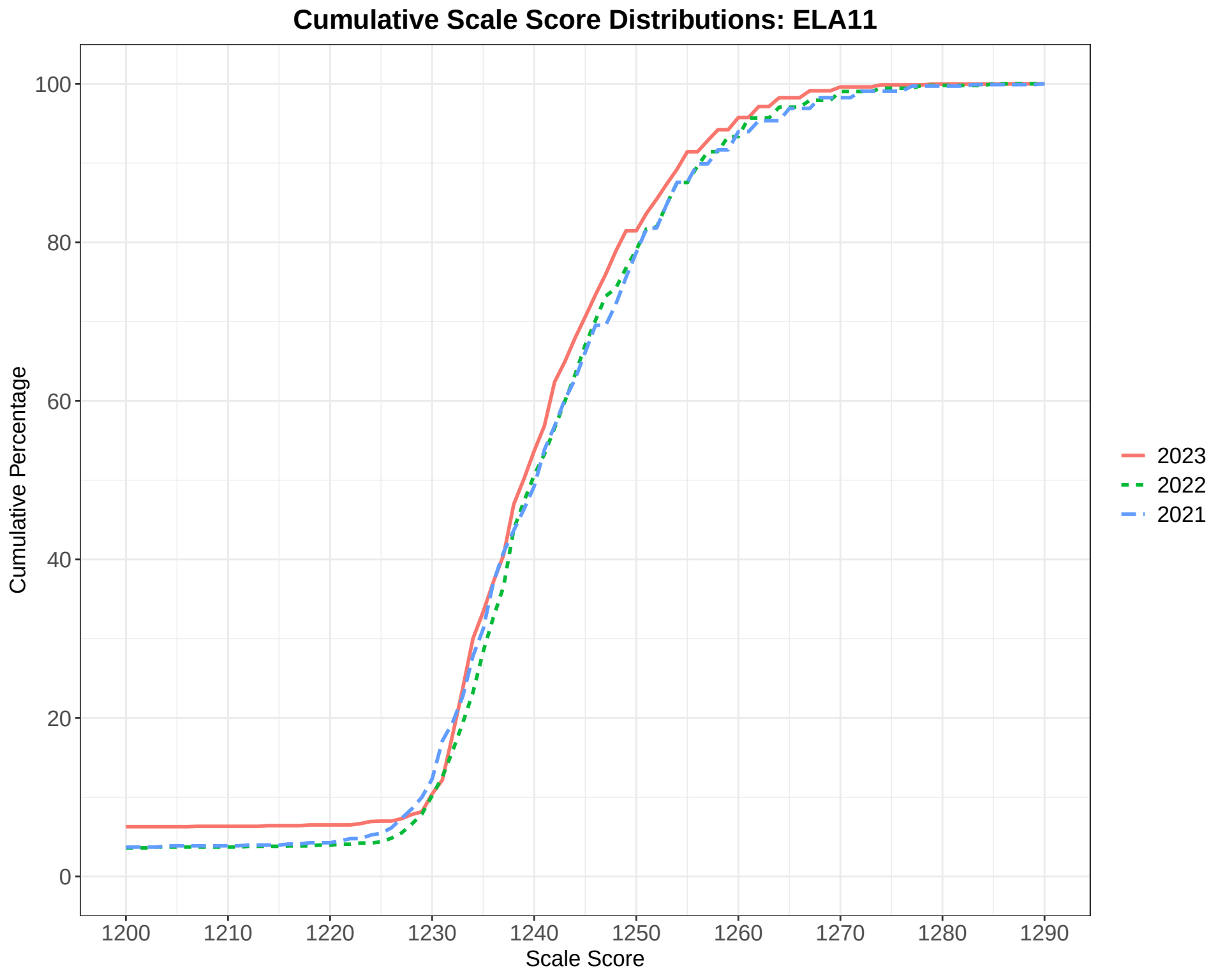

Cumulative Scale Score Distributions: ELA04
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021

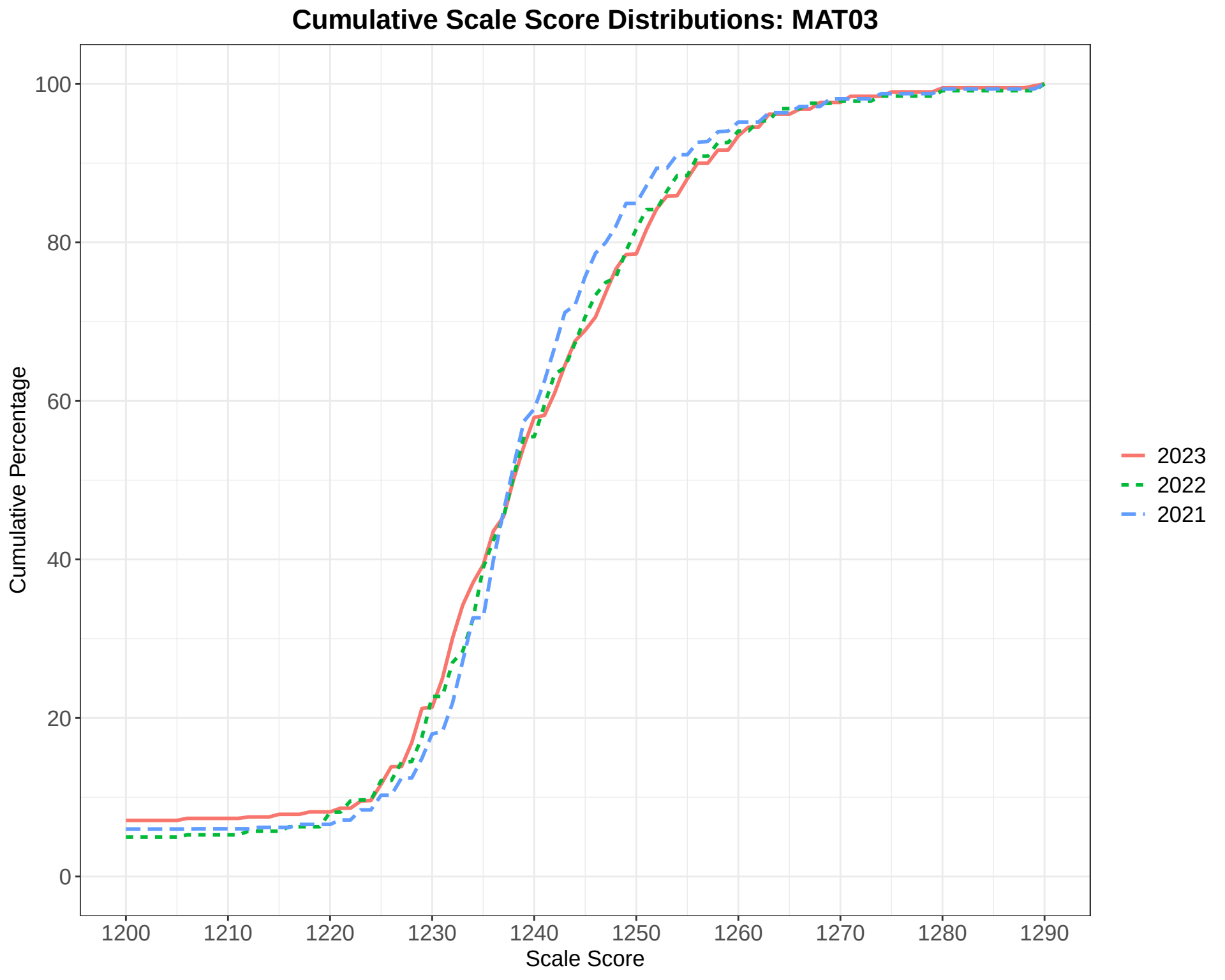

Cumulative Scale Score Distributions: ELA05
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021

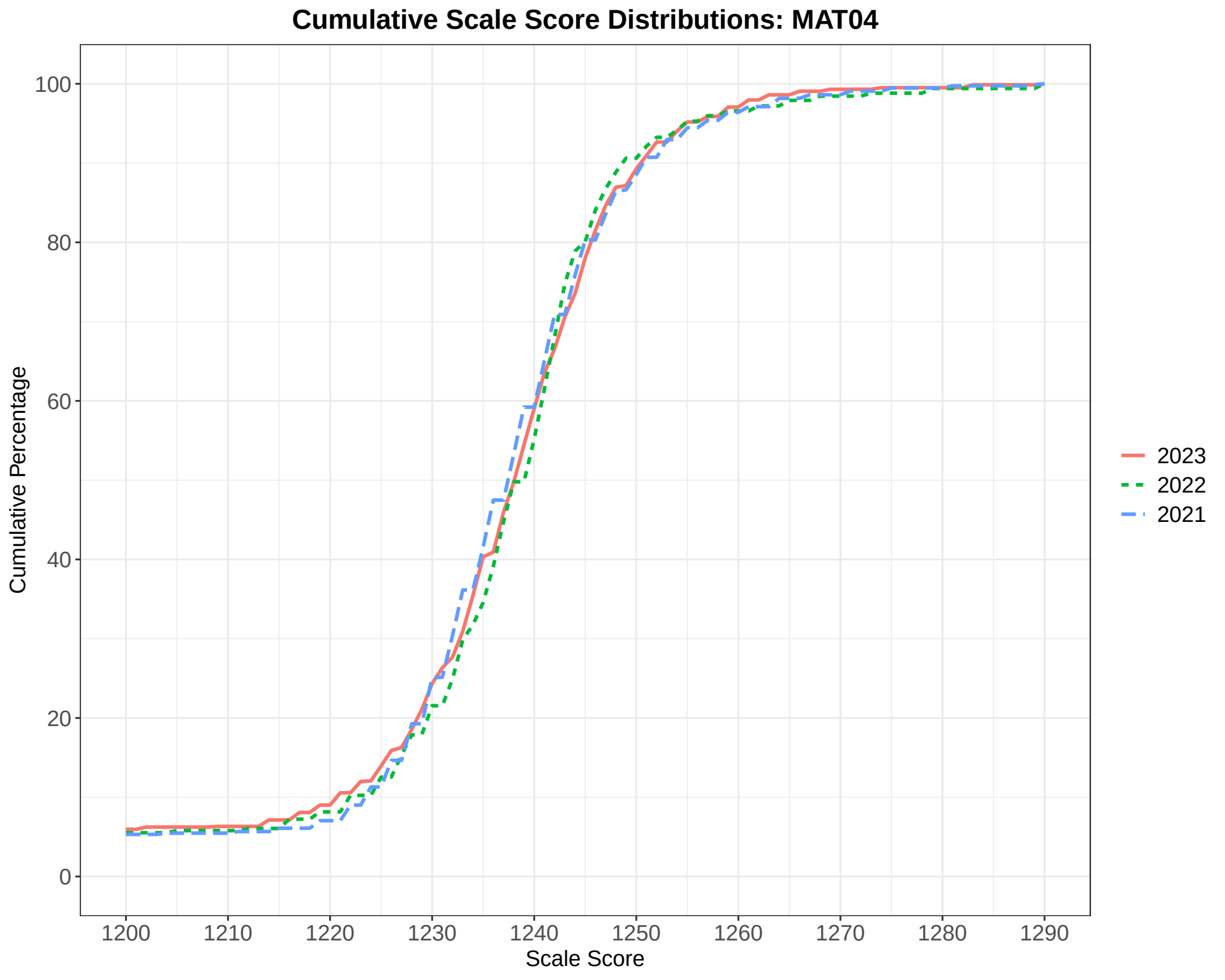

Cumulative Scale Score Distributions: ELA06
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021

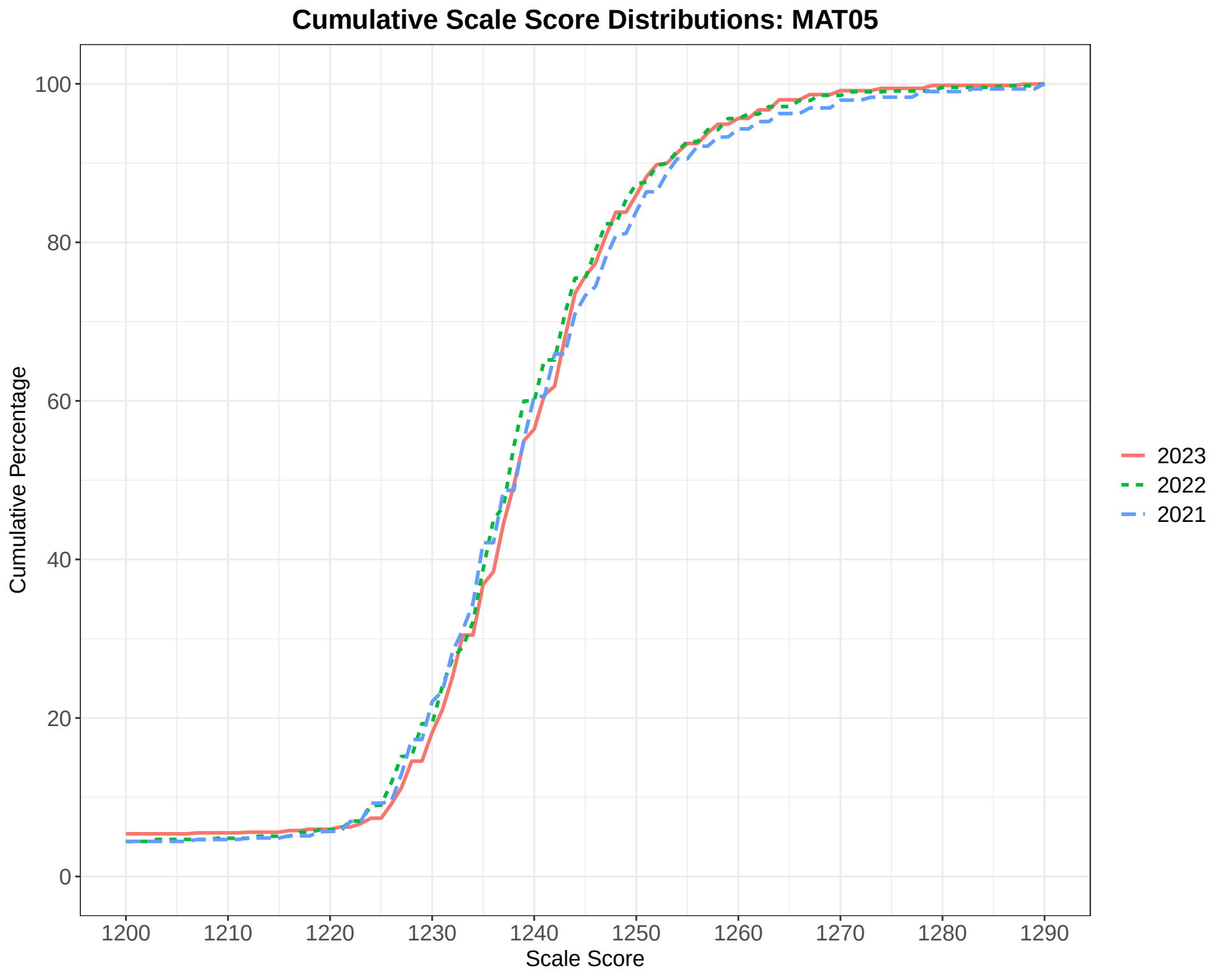

Cumulative Scale Score Distributions: ELA07
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021

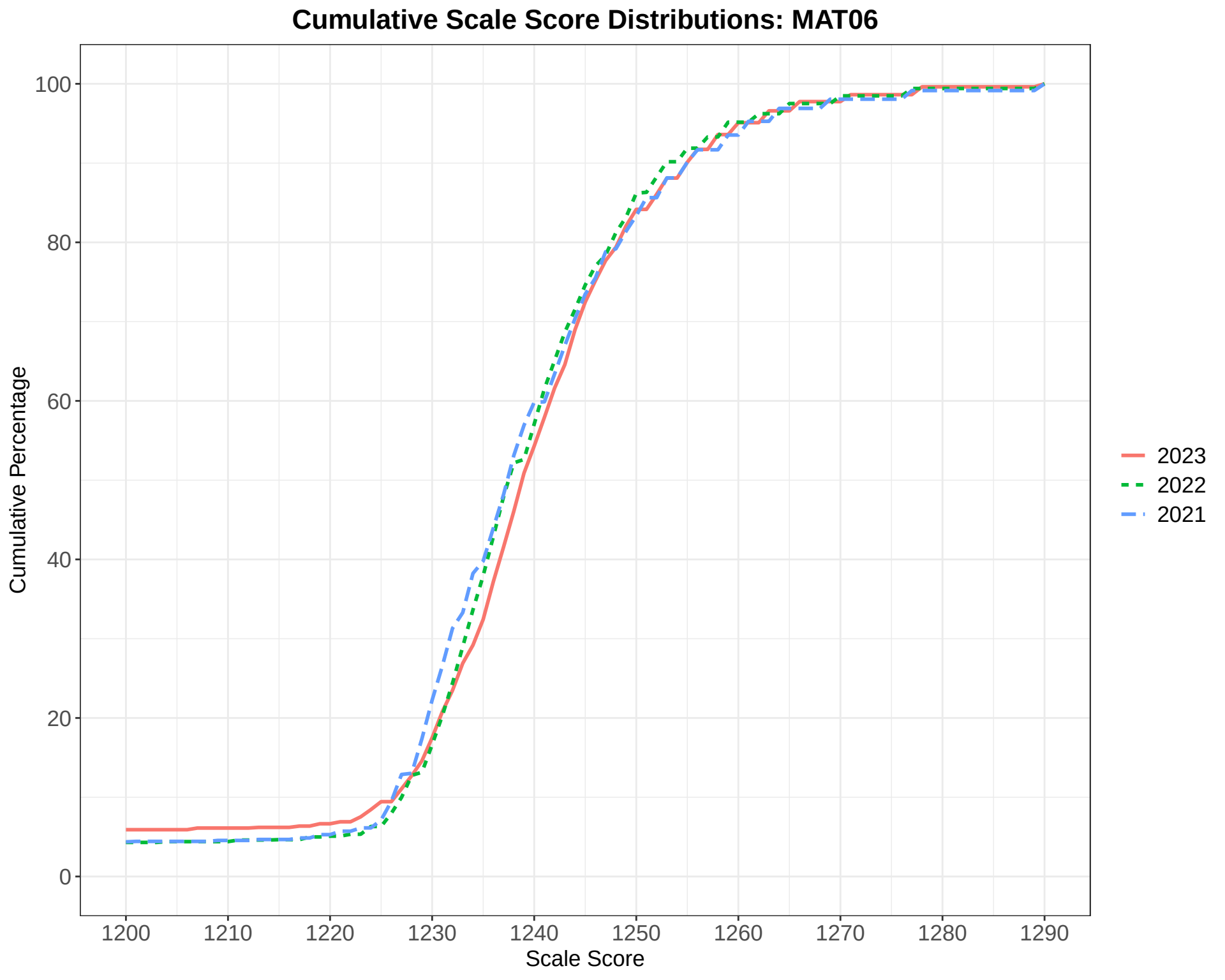

Cumulative Scale Score Distributions: ELA08
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021

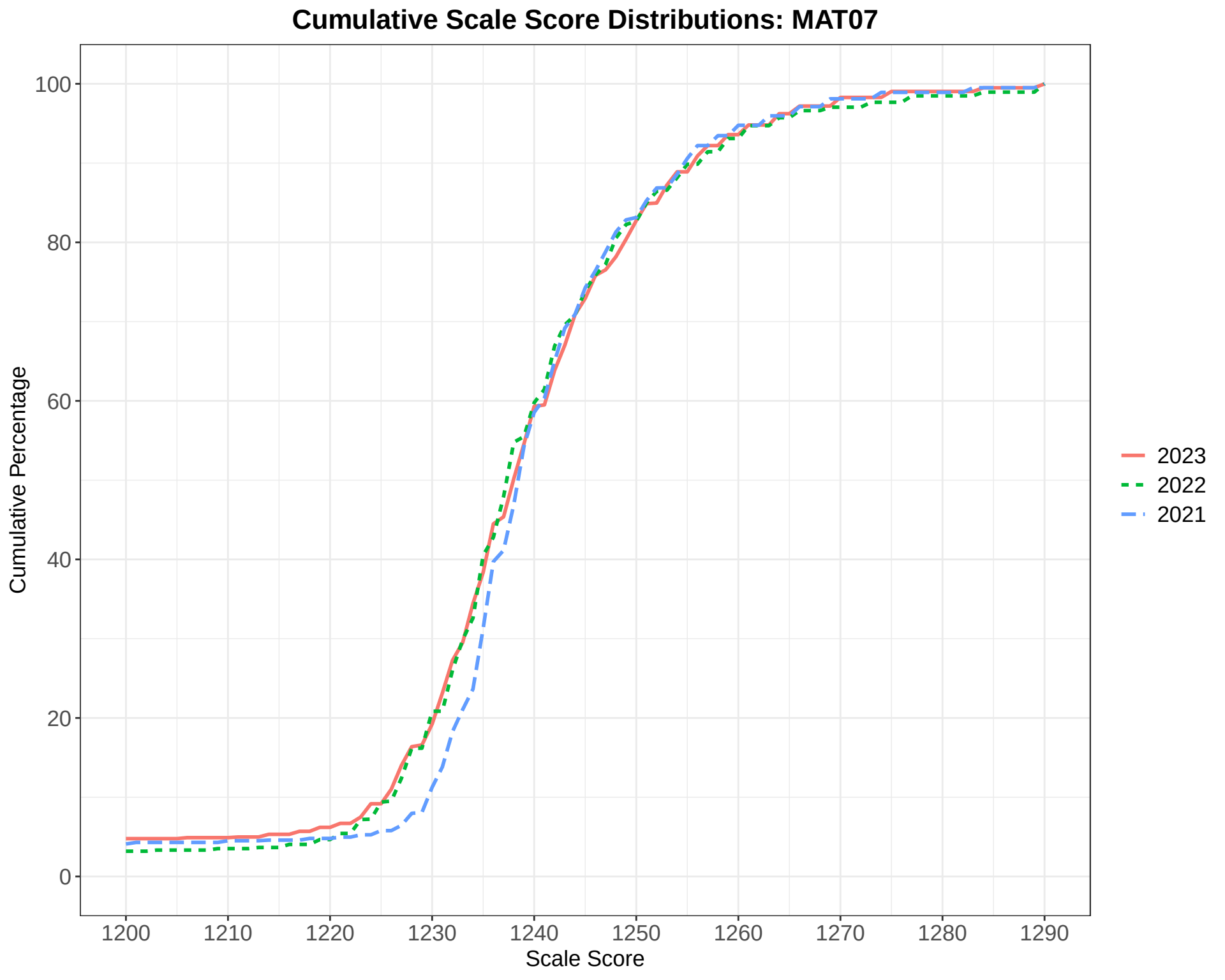

Cumulative Scale Score Distributions: ELA11
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021

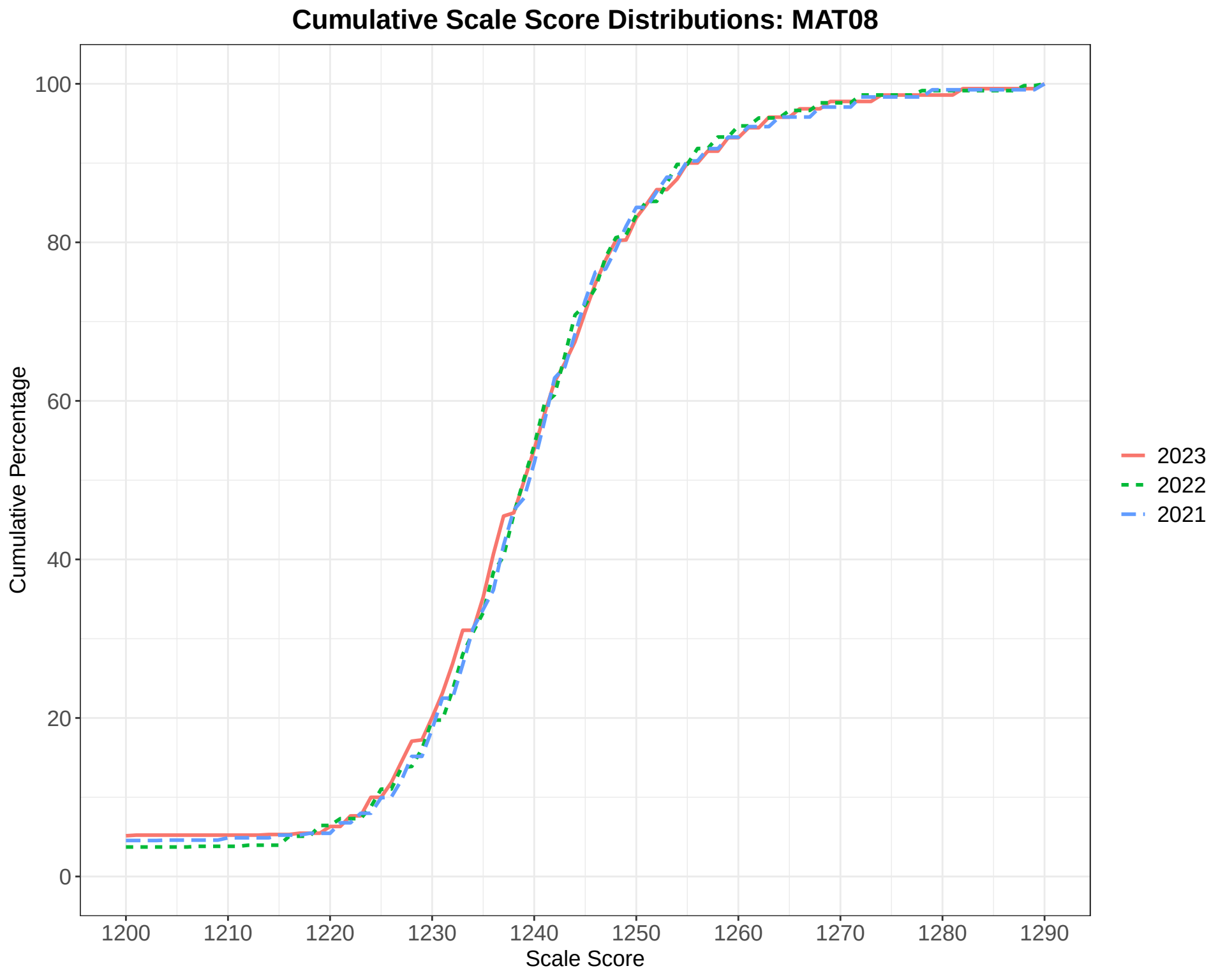

Cumulative Scale Score Distributions: MAT03
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021

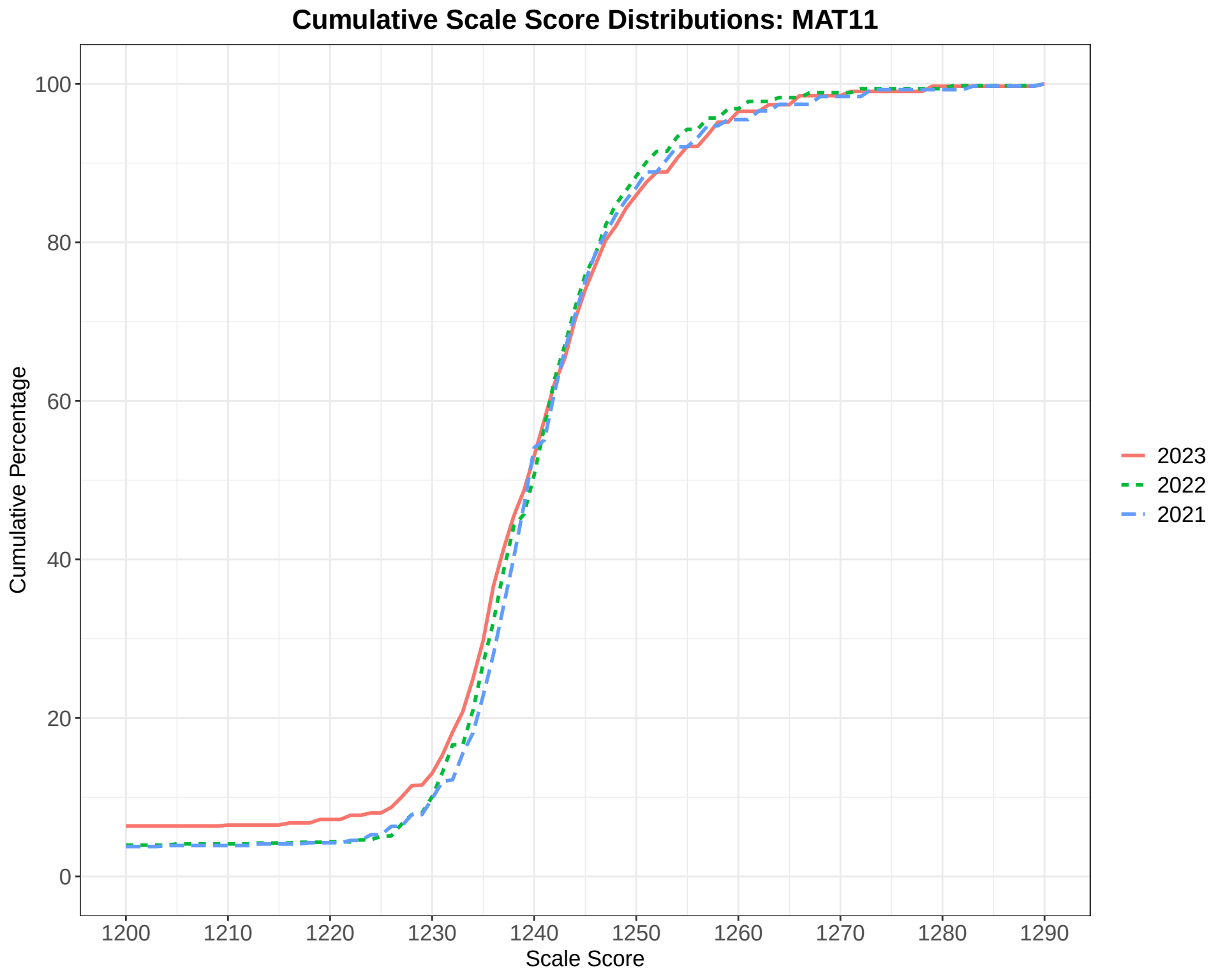

Cumulative Scale Score Distributions: MAT04
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021


Cumulative Scale Score Distributions: MAT05
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021


Cumulative Scale Score Distributions: MAT06
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021


Cumulative Scale Score Distributions: MAT07
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021


Cumulative Scale Score Distributions: MAT08
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021


Cumulative Scale Score Distributions: MAT11
Cumulative Percentage
100
80
60
40
20
0
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
Scale Score
2023
2022
2021

APPENDIX M
IRT SUBGROUP RELIABILITY

Note: Values are calculated only for subgroups with 100 or more students.

Table M-1. IRT Subgroup Reliability: ELA Grade 3

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	5,559	1200	1289	1240.31	11.17	0.92	3.04
Female	1,737	1200	1289	1240.15	11.30	0.92	3.04
Male	3,783	1200	1281	1240.37	11.11	0.92	3.04
Gender Undefined	39	1228	1266	1242.23	11.50	NA	NA
Hispanic or Latino	1,308	1200	1275	1238.34	10.46	0.92	2.88
American Indian or Alaska Native	228	1206	1263	1239.37	10.49	0.91	2.96
Asian	195	1224	1266	1239.02	8.91	0.89	2.81
Black or African American	948	1217	1289	1243.59	13.27	0.92	3.39
Native Hawaiian or Pacific Islander	51	1219	1248	1234.06	7.40	NA	NA
White (non-Hispanic)	2,340	1200	1281	1240.06	10.59	0.91	3.00
Two or More Races (non-Hispanic)	237	1223	1263	1239.70	9.78	0.91	2.91
No Primary race/Ethnicity Undefined	1,308	1200	1275	1238.34	10.46	0.92	2.88
Currently receiving LEP services	666	1200	1275	1237.32	9.81	0.91	2.81
Not receiving LEP services	3,039	1200	1289	1241.37	11.55	0.92	3.13
Economically Disadvantaged Students	1,629	1206	1281	1242.79	11.87	0.91	3.24
Non-economically Disadvantaged Students	1,674	1200	1289	1239.82	10.95	0.91	3.00
Migrant	75	1224	1266	1237.52	9.85	NA	NA
Non-migrant	3,171	1200	1289	1241.32	11.57	0.92	3.13
Augmentative Communication	1,200	1200	1270	1234.02	8.17	0.89	2.64
No Augmentative Communication	4,338	1200	1289	1242.07	11.27	0.91	3.15
Undefined Augmentative Communications	21	1228	1250	1236.43	8.57	NA	NA
Hearing Loss	78	1211	1246	1230.15	6.36	NA	NA
Within Normal Limits	5,481	1200	1289	1240.46	11.16	0.92	3.05
Visual Impairment	225	1200	1275	1236.32	12.10	0.92	3.07
Within Normal Limits	5,316	1200	1289	1240.49	11.10	0.92	3.04
Undefined Visual Impairment	18	1227	1263	1239.33	12.71	NA	NA
Sensory Stimuli Response	510	1200	1263	1230.15	7.98	0.86	2.73
Follow Directions	5,049	1200	1289	1241.34	10.93	0.91	3.07
Special School	267	1200	1275	1234.83	10.94	0.92	2.87
Regular School Self-contained	3,591	1200	1289	1239.47	11.15	0.92	2.99
Regular School Primarily Self-contained	1,017	1216	1281	1242.25	10.69	0.91	3.12
Regular School Resource Room	459	1219	1275	1243.73	10.64	0.90	3.23
Regular School General Education	225	1227	1275	1244.57	10.37	0.89	3.27
Communicates Primarily Through Cries	411	1200	1260	1230.19	8.79	0.88	2.79
Uses Intentional Communication	1,545	1200	1275	1235.23	8.99	0.90	2.70
Uses Symbolic Language	3,603	1213	1289	1243.65	10.74	0.90	3.21

Table M-2. IRT Subgroup Reliability: ELA Grade 4

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	5,778	1200	1290	1240.39	12.37	0.93	3.16
Female	1,920	1200	1288	1240.43	12.47	0.93	3.18
Male	3,834	1200	1290	1240.36	12.32	0.93	3.15
Gender Undefined	24	1228	1265	1242.00	14.41	NA	NA
Hispanic or Latino	1,368	1200	1290	1238.86	12.13	0.92	3.10
American Indian or Alaska Native	243	1203	1268	1239.04	11.69	0.92	3.10
Asian	186	1215	1268	1236.48	10.17	0.92	2.88
Black or African American	1,011	1207	1282	1241.51	12.82	0.93	3.23
Native Hawaiian or Pacific Islander	51	1215	1251	1233.35	9.11	NA	NA
White (non-Hispanic)	2,412	1203	1290	1241.14	12.42	0.93	3.19
Two or More Races (non-Hispanic)	294	1215	1276	1240.12	11.93	0.93	3.11
No Primary race/Ethnicity Undefined	1,368	1200	1290	1238.86	12.13	0.92	3.10
Currently receiving LEP services	462	1215	1268	1238.09	10.63	0.92	2.95
Not receiving LEP services	3,315	1203	1290	1241.52	12.55	0.93	3.22
Economically Disadvantaged Students	1,764	1203	1290	1242.63	12.65	0.92	3.30
Non-economically Disadvantaged Students	1,746	1215	1290	1240.00	12.16	0.93	3.11
Migrant	153	1215	1263	1239.39	11.82	0.93	3.05
Non-migrant	3,321	1203	1290	1241.42	12.53	0.93	3.21
Augmentative Communication	1,068	1200	1276	1232.48	8.52	0.89	2.70
No Augmentative Communication	4,665	1200	1290	1242.22	12.38	0.92	3.27
Undefined Augmentative Communications	45	1221	1272	1238.33	14.76	NA	NA
Hearing Loss	126	1200	1276	1234.07	13.12	0.90	3.35
Within Normal Limits	5,652	1200	1290	1240.53	12.32	0.93	3.16
Visual Impairment	171	1215	1290	1235.18	12.32	0.93	2.93
Within Normal Limits	5,592	1200	1290	1240.57	12.35	0.93	3.17
Undefined Visual Impairment	15	1224	1244	1232.80	7.03	NA	NA
Sensory Stimuli Response	387	1200	1254	1228.19	8.97	0.85	2.93
Follow Directions	5,391	1203	1290	1241.27	12.12	0.92	3.18
Special School	273	1207	1255	1233.16	9.33	0.91	2.76
Regular School Self-contained	3,648	1200	1290	1239.05	12.15	0.93	3.09
Regular School Primarily Self-contained	1,161	1215	1276	1243.30	12.18	0.92	3.31
Regular School Resource Room	438	1211	1276	1245.98	12.35	0.91	3.51
Regular School General Education	258	1223	1272	1244.38	11.71	0.91	3.34
Communicates Primarily Through Cries	318	1200	1276	1229.24	8.52	0.88	2.75
Uses Intentional Communication	1,359	1200	1282	1234.13	10.09	0.91	2.83
Uses Symbolic Language	4,101	1203	1290	1243.33	12.05	0.92	3.30

Table M-3. IRT Subgroup Reliability: ELA Grade 5

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	5,799	1207	1290	1239.74	11.73	0.92	3.06
Female	1,920	1211	1281	1240.20	11.78	0.92	3.09
Male	3,858	1207	1290	1239.46	11.66	0.92	3.04
Gender Undefined	21	1228	1268	1249.29	14.50	NA	NA
Hispanic or Latino	1,338	1219	1281	1237.67	10.55	0.92	2.90
American Indian or Alaska Native	213	1225	1261	1238.52	9.70	0.91	2.87
Asian	150	1211	1276	1233.42	9.51	0.91	2.68
Black or African American	1,149	1217	1290	1241.80	12.20	0.92	3.22
Native Hawaiian or Pacific Islander	48	1211	1259	1232.31	10.78	NA	NA
White (non-Hispanic)	2,457	1207	1281	1240.28	11.84	0.93	3.09
Two or More Races (non-Hispanic)	255	1214	1268	1238.79	11.51	0.93	3.02
No Primary race/Ethnicity Undefined	1,338	1219	1281	1237.67	10.55	0.92	2.90
Currently receiving LEP services	414	1221	1261	1237.49	9.04	0.89	2.87
Not receiving LEP services	3,405	1207	1290	1240.74	11.91	0.92	3.13
Economically Disadvantaged Students	1,719	1214	1281	1242.22	12.06	0.92	3.24
Non-economically Disadvantaged Students	1,848	1207	1290	1238.97	11.12	0.92	2.98
Migrant	150	1225	1263	1235.36	8.41	0.89	2.65
Non-migrant	3,369	1207	1290	1240.73	11.81	0.92	3.12
Augmentative Communication	993	1211	1271	1231.82	7.69	0.89	2.50
No Augmentative Communication	4,767	1207	1290	1241.33	11.75	0.92	3.17
Undefined Augmentative Communications	39	1228	1263	1246.85	10.68	NA	NA
Hearing Loss	114	1216	1261	1233.95	9.72	0.92	2.64
Within Normal Limits	5,685	1207	1290	1239.86	11.74	0.92	3.07
Visual Impairment	204	1214	1281	1234.21	11.05	0.93	2.82
Within Normal Limits	5,553	1207	1290	1239.95	11.70	0.92	3.07
Undefined Visual Impairment	42	1225	1263	1238.21	11.60	NA	NA
Sensory Stimuli Response	336	1216	1252	1230.12	6.57	0.86	2.43
Follow Directions	5,463	1207	1290	1240.33	11.72	0.92	3.10
Special School	300	1216	1265	1231.99	8.46	0.91	2.50
Regular School Self-contained	3,480	1211	1290	1238.77	11.49	0.93	2.99
Regular School Primarily Self-contained	1,278	1220	1281	1242.05	11.57	0.92	3.20
Regular School Resource Room	441	1207	1276	1241.80	11.94	0.92	3.27
Regular School General Education	300	1225	1281	1245.88	11.77	0.90	3.52
Communicates Primarily Through Cries	288	1216	1252	1229.39	5.91	0.83	2.41
Uses Intentional Communication	1,239	1211	1276	1233.64	9.42	0.91	2.65
Uses Symbolic Language	4,272	1207	1290	1242.21	11.60	0.92	3.22

Table M-4. IRT Subgroup Reliability: ELA Grade 6

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	5,721	1203	1290	1239.20	9.70	0.90	2.91
Female	1,899	1203	1290	1239.49	9.66	0.89	2.94
Male	3,798	1203	1290	1239.08	9.74	0.90	2.90
Gender Undefined	24	1229	1243	1236.50	4.57	NA	NA
Hispanic or Latino	1,377	1203	1277	1238.48	9.47	0.90	2.85
American Indian or Alaska Native	261	1225	1260	1238.75	7.79	0.86	2.79
Asian	114	1226	1266	1236.84	8.84	0.90	2.62
Black or African American	1,137	1203	1277	1239.00	9.82	0.90	2.93
Native Hawaiian or Pacific Islander	45	1232	1255	1238.93	6.54	NA	NA
White (non-Hispanic)	2,346	1209	1277	1239.58	9.79	0.90	2.93
Two or More Races (non-Hispanic)	264	1220	1290	1239.89	10.21	0.89	3.02
No Primary race/Ethnicity Undefined	1,377	1203	1277	1238.48	9.47	0.90	2.85
Currently receiving LEP services	387	1209	1263	1238.02	9.02	0.90	2.79
Not receiving LEP services	3,294	1203	1277	1239.36	9.60	0.90	2.92
Economically Disadvantaged Students	1,620	1203	1277	1239.80	9.52	0.89	2.96
Non-economically Disadvantaged Students	1,833	1213	1271	1238.69	9.65	0.90	2.85
Migrant	153	1203	1260	1238.08	9.79	0.90	2.96
Non-migrant	3,264	1209	1277	1239.23	9.60	0.90	2.90
Augmentative Communication	855	1203	1263	1233.14	7.39	0.88	2.46
No Augmentative Communication	4,818	1203	1290	1240.27	9.67	0.89	2.99
Undefined Augmentative Communications	48	1226	1263	1240.44	9.53	NA	NA
Hearing Loss	90	1226	1257	1235.37	8.33	NA	NA
Within Normal Limits	5,631	1203	1290	1239.27	9.71	0.90	2.92
Visual Impairment	204	1209	1266	1234.66	9.13	0.91	2.58
Within Normal Limits	5,484	1203	1290	1239.38	9.68	0.90	2.92
Undefined Visual Impairment	33	1220	1257	1238.64	10.34	NA	NA
Sensory Stimuli Response	330	1209	1255	1231.69	7.75	0.89	2.43
Follow Directions	5,391	1203	1290	1239.66	9.62	0.89	2.94
Special School	282	1203	1263	1233.56	10.07	0.92	2.65
Regular School Self-contained	3,726	1203	1277	1238.29	9.17	0.89	2.81
Regular School Primarily Self-contained	1,116	1222	1290	1240.86	9.86	0.89	3.03
Regular School Resource Room	366	1227	1277	1244.70	10.24	0.87	3.45
Regular School General Education	231	1226	1263	1244.08	8.86	0.85	3.32
Communicates Primarily Through Cries	324	1209	1251	1231.61	6.72	0.87	2.35
Uses Intentional Communication	1,107	1203	1263	1234.83	8.33	0.89	2.59
Uses Symbolic Language	4,290	1209	1290	1240.91	9.59	0.89	3.03

Table M-5. IRT Subgroup Reliability: ELA Grade 7

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	5,832	1203	1290	1242.67	9.81	0.90	3.04
Female	2,085	1203	1272	1242.77	9.85	0.90	3.06
Male	3,720	1203	1290	1242.60	9.80	0.90	3.03
Gender Undefined	27	1236	1262	1245.00	8.53	NA	NA
Hispanic or Latino	1,392	1220	1268	1241.46	9.37	0.90	2.94
American Indian or Alaska Native	228	1220	1262	1240.99	9.65	0.90	2.94
Asian	114	1219	1262	1238.61	10.00	0.91	2.92
Black or African American	1,152	1203	1272	1243.96	10.03	0.89	3.16
Native Hawaiian or Pacific Islander	60	1224	1254	1238.10	8.28	NA	NA
White (non-Hispanic)	2,463	1203	1290	1243.08	9.79	0.90	3.06
Two or More Races (non-Hispanic)	243	1220	1268	1241.96	9.61	0.90	2.98
No Primary race/Ethnicity Undefined	1,392	1220	1268	1241.46	9.37	0.90	2.94
Currently receiving LEP services	315	1203	1268	1240.40	9.94	0.90	2.97
Not receiving LEP services	3,429	1203	1278	1243.47	9.70	0.89	3.09
Economically Disadvantaged Students	1,662	1203	1278	1244.22	9.88	0.89	3.16
Non-economically Disadvantaged Students	1,923	1203	1268	1242.50	9.52	0.89	3.01
Migrant	180	1203	1268	1241.18	11.08	0.91	3.16
Non-migrant	3,360	1203	1278	1243.44	9.63	0.89	3.08
Augmentative Communication	786	1203	1262	1235.50	7.96	0.88	2.69
No Augmentative Communication	5,007	1203	1290	1243.78	9.63	0.89	3.10
Undefined Augmentative Communications	39	1234	1249	1243.77	4.25	NA	NA
Hearing Loss	117	1220	1256	1236.21	9.19	0.91	2.68
Within Normal Limits	5,715	1203	1290	1242.80	9.78	0.90	3.05
Visual Impairment	204	1220	1254	1236.41	8.13	0.89	2.68
Within Normal Limits	5,601	1203	1290	1242.88	9.78	0.90	3.05
Undefined Visual Impairment	27	1226	1265	1246.22	12.35	NA	NA
Sensory Stimuli Response	312	1203	1262	1233.53	8.34	0.88	2.74
Follow Directions	5,520	1203	1290	1243.19	9.64	0.89	3.06
Special School	279	1203	1262	1236.33	9.13	0.89	2.84
Regular School Self-contained	3,894	1203	1290	1241.84	9.67	0.90	2.99
Regular School Primarily Self-contained	1,044	1220	1278	1245.05	9.63	0.88	3.17
Regular School Resource Room	384	1224	1272	1246.45	8.92	0.86	3.24
Regular School General Education	228	1222	1265	1247.11	8.62	0.85	3.28
Communicates Primarily Through Cries	279	1203	1254	1232.43	7.51	0.86	2.71
Uses Intentional Communication	1,071	1222	1265	1238.25	8.67	0.89	2.77
Uses Symbolic Language	4,482	1203	1290	1244.36	9.48	0.88	3.13

Table M-6. IRT Subgroup Reliability: ELA Grade 8

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	5,868	1200	1290	1238.49	10.26	0.91	2.92
Female	2,169	1211	1290	1238.54	9.98	0.90	2.91
Male	3,672	1200	1290	1238.50	10.43	0.91	2.92
Gender Undefined	27	1224	1250	1233.56	8.39	NA	NA
Hispanic or Latino	1,296	1200	1290	1236.98	10.01	0.91	2.81
American Indian or Alaska Native	246	1200	1257	1234.56	9.24	0.90	2.71
Asian	126	1220	1255	1235.45	8.59	0.90	2.59
Black or African American	1,155	1200	1290	1239.97	10.52	0.90	3.07
Native Hawaiian or Pacific Islander	63	1211	1266	1234.38	10.29	NA	NA
White (non-Hispanic)	2,622	1216	1290	1239.09	10.13	0.90	2.93
Two or More Races (non-Hispanic)	207	1217	1271	1238.68	10.58	0.91	2.92
No Primary race/Ethnicity Undefined	1,296	1200	1290	1236.98	10.01	0.91	2.81
Currently receiving LEP services	291	1200	1271	1235.95	9.71	0.90	2.77
Not receiving LEP services	3,600	1200	1290	1238.96	10.39	0.91	2.96
Economically Disadvantaged Students	1,737	1200	1290	1239.21	10.54	0.90	3.01
Non-economically Disadvantaged Students	1,989	1216	1279	1238.59	10.12	0.91	2.89
Migrant	201	1200	1266	1235.03	10.99	0.92	2.86
Non-migrant	3,471	1200	1290	1239.15	10.27	0.90	2.96
Augmentative Communication	750	1211	1266	1231.76	7.73	0.89	2.44
No Augmentative Communication	5,061	1200	1290	1239.49	10.23	0.90	2.99
Undefined Augmentative Communications	57	1226	1255	1238.74	8.73	NA	NA
Hearing Loss	117	1200	1260	1235.15	11.28	0.92	2.90
Within Normal Limits	5,751	1200	1290	1238.56	10.23	0.91	2.92
Visual Impairment	189	1200	1260	1232.48	9.92	0.91	2.67
Within Normal Limits	5,658	1200	1290	1238.71	10.21	0.91	2.93
Undefined Visual Impairment	21	1224	1248	1234.43	8.81	NA	NA
Sensory Stimuli Response	249	1200	1253	1229.33	7.36	0.87	2.45
Follow Directions	5,619	1200	1290	1238.90	10.18	0.90	2.94
Special School	327	1200	1260	1230.28	8.35	0.90	2.51
Regular School Self-contained	3,702	1200	1290	1237.55	9.91	0.91	2.83
Regular School Primarily Self-contained	1,206	1211	1290	1241.45	9.84	0.88	3.12
Regular School Resource Room	405	1224	1271	1242.86	10.65	0.90	3.25
Regular School General Education	228	1224	1266	1242.25	9.86	0.89	3.15
Communicates Primarily Through Cries	216	1200	1245	1227.86	7.23	0.86	2.47
Uses Intentional Communication	1,014	1200	1290	1233.05	8.80	0.90	2.55
Uses Symbolic Language	4,638	1200	1290	1240.18	9.96	0.90	3.02

Table M-7. IRT Subgroup Reliability: ELA Grade HS

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	5,586	1207	1287	1243.40	9.51	0.91	2.65
Female	1,947	1218	1279	1243.85	9.52	0.91	2.68
Male	3,525	1207	1287	1243.12	9.51	0.92	2.64
Gender Undefined	114	1227	1270	1244.37	9.36	0.91	2.71
Hispanic or Latino	1,131	1214	1270	1242.21	8.67	0.91	2.53
American Indian or Alaska Native	168	1224	1264	1242.59	8.26	0.90	2.61
Asian	105	1218	1260	1238.26	7.68	0.90	2.35
Black or African American	1,059	1207	1287	1243.25	9.61	0.91	2.67
Native Hawaiian or Pacific Islander	24	1232	1252	1239.50	6.00	NA	NA
White (non-Hispanic)	2,640	1218	1274	1244.04	9.61	0.91	2.70
Two or More Races (non-Hispanic)	234	1227	1279	1243.47	11.32	0.93	2.71
No Primary race/Ethnicity Undefined	1,131	1214	1270	1242.21	8.67	0.91	2.53
Currently receiving LEP services	213	1223	1267	1240.06	8.83	0.92	2.42
Not receiving LEP services	3,483	1207	1287	1244.00	9.52	0.91	2.70
Economically Disadvantaged Students	1,461	1224	1287	1244.68	9.51	0.91	2.74
Non-economically Disadvantaged Students	2,169	1207	1274	1243.33	9.43	0.91	2.65
Migrant	102	1223	1262	1240.62	8.58	0.91	2.49
Non-migrant	3,513	1207	1287	1243.97	9.49	0.91	2.69
Augmentative Communication	612	1214	1264	1237.31	8.54	0.92	2.25
No Augmentative Communication	4,947	1207	1287	1244.15	9.36	0.91	2.71
Undefined Augmentative Communications	27	1233	1257	1243.44	8.29	NA	NA
Hearing Loss	135	1227	1257	1238.44	6.75	0.89	2.23
Within Normal Limits	5,451	1207	1287	1243.52	9.54	0.91	2.67
Visual Impairment	153	1207	1274	1241.10	12.32	0.94	2.78
Within Normal Limits	5,415	1214	1287	1243.48	9.42	0.91	2.65
Undefined Visual Impairment	18	1232	1243	1237.17	4.66	NA	NA
Sensory Stimuli Response	207	1207	1260	1232.90	6.72	0.86	2.25
Follow Directions	5,379	1214	1287	1243.80	9.37	0.91	2.67
Special School	366	1214	1274	1238.58	9.99	0.93	2.40
Regular School Self-contained	3,519	1207	1287	1242.56	9.35	0.91	2.59
Regular School Primarily Self-contained	1,155	1223	1274	1245.35	8.89	0.90	2.77
Regular School Resource Room	423	1228	1274	1247.07	8.70	0.88	2.91
Regular School General Education	123	1237	1274	1250.78	9.38	0.87	3.25
Communicates Primarily Through Cries	207	1207	1260	1233.99	7.21	0.88	2.27
Uses Intentional Communication	909	1214	1270	1237.79	7.92	0.91	2.26
Uses Symbolic Language	4,470	1218	1287	1244.98	9.22	0.90	2.75

Table M-8. IRT Subgroup Reliability: Mathematics Grade 3

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	5,673	1200	1290	1242.41	10.86	0.87	3.76
Female	1,755	1206	1290	1242.21	10.87	0.86	3.76
Male	3,873	1200	1290	1242.49	10.88	0.87	3.76
Gender Undefined	45	1229	1254	1243.07	8.12	NA	NA
Hispanic or Latino	1,290	1200	1290	1241.20	10.26	0.86	3.68
American Indian or Alaska Native	231	1206	1274	1243.29	10.26	0.86	3.70
Asian	195	1212	1274	1241.46	11.38	0.88	3.76
Black or African American	981	1215	1290	1245.46	12.41	0.86	4.08
Native Hawaiian or Pacific Islander	42	1228	1254	1240.07	8.14	NA	NA
White (non-Hispanic)	2,421	1200	1282	1241.66	10.35	0.87	3.69
Two or More Races (non-Hispanic)	255	1206	1258	1240.40	9.71	0.86	3.59
No Primary race/Ethnicity Undefined	1,290	1200	1290	1241.20	10.26	0.86	3.68
Currently receiving LEP services	660	1200	1269	1240.50	10.65	0.88	3.67
Not receiving LEP services	3,126	1200	1290	1242.73	10.88	0.86	3.78
Economically Disadvantaged Students	1,638	1215	1290	1244.25	10.76	0.85	3.82
Non-economically Disadvantaged Students	1,767	1200	1290	1241.38	10.78	0.87	3.72
Migrant	96	1225	1269	1240.47	10.81	NA	NA
Non-migrant	3,246	1200	1290	1242.78	10.88	0.86	3.77
Augmentative Communication	1,248	1206	1262	1237.08	9.13	0.84	3.59
No Augmentative Communication	4,401	1200	1290	1243.94	10.83	0.86	3.81
Undefined Augmentative Communications	24	1215	1258	1239.25	11.81	NA	NA
Hearing Loss	75	1206	1247	1234.28	9.53	NA	NA
Within Normal Limits	5,598	1200	1290	1242.52	10.84	0.86	3.76
Visual Impairment	216	1206	1274	1237.83	11.93	0.89	3.83
Within Normal Limits	5,439	1200	1290	1242.60	10.78	0.86	3.76
Undefined Visual Impairment	18	1229	1250	1238.33	8.60	NA	NA
Sensory Stimuli Response	510	1200	1269	1234.59	9.95	0.85	3.73
Follow Directions	5,163	1200	1290	1243.18	10.64	0.86	3.77
Special School	273	1206	1265	1237.62	9.72	0.86	3.60
Regular School Self-contained	3,672	1200	1290	1241.86	11.03	0.87	3.77
Regular School Primarily Self-contained	1,071	1206	1290	1243.28	10.26	0.86	3.73
Regular School Resource Room	441	1225	1290	1245.49	9.73	0.82	3.81
Regular School General Education	216	1228	1290	1247.19	10.86	0.83	4.01
Communicates Primarily Through Cries	396	1200	1269	1234.12	10.66	0.86	3.83
Uses Intentional Communication	1,620	1206	1282	1238.99	10.46	0.87	3.67
Uses Symbolic Language	3,657	1200	1290	1244.82	10.19	0.84	3.80

Table M-9. IRT Subgroup Reliability: Mathematics Grade 4

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	5,979	1203	1290	1239.02	10.02	0.88	3.37
Female	1,998	1203	1274	1238.01	9.43	0.87	3.33
Male	3,960	1203	1290	1239.53	10.26	0.88	3.39
Gender Undefined	21	1228	1263	1240.00	11.05	NA	NA
Hispanic or Latino	1,398	1203	1266	1238.59	8.76	0.86	3.26
American Indian or Alaska Native	246	1225	1261	1238.32	7.31	0.81	3.17
Asian	195	1219	1283	1236.55	10.54	0.88	3.46
Black or African American	1,056	1203	1290	1239.39	10.21	0.88	3.38
Native Hawaiian or Pacific Islander	57	1222	1257	1233.63	7.73	NA	NA
White (non-Hispanic)	2,502	1203	1290	1239.26	10.32	0.88	3.41
Two or More Races (non-Hispanic)	315	1216	1283	1237.96	10.52	0.89	3.42
No Primary race/Ethnicity Undefined	1,398	1203	1266	1238.59	8.76	0.86	3.26
Currently receiving LEP services	480	1219	1283	1238.34	9.71	0.88	3.33
Not receiving LEP services	3,402	1203	1290	1239.61	10.14	0.88	3.37
Economically Disadvantaged Students	1,827	1203	1283	1240.38	10.04	0.88	3.35
Non-economically Disadvantaged Students	1,776	1216	1290	1238.66	9.96	0.87	3.38
Migrant	162	1219	1261	1237.74	9.11	0.87	3.29
Non-migrant	3,405	1203	1290	1239.62	10.10	0.88	3.37
Augmentative Communication	1,212	1203	1274	1234.34	7.86	0.80	3.39
No Augmentative Communication	4,719	1203	1290	1240.21	10.16	0.88	3.36
Undefined Augmentative Communications	48	1227	1263	1240.50	10.27	NA	NA
Hearing Loss	120	1216	1259	1234.22	8.31	0.83	3.42
Within Normal Limits	5,859	1203	1290	1239.12	10.03	0.88	3.37
Visual Impairment	195	1203	1259	1234.22	9.68	0.85	3.54
Within Normal Limits	5,772	1203	1290	1239.19	10.00	0.88	3.36
Sensory Stimuli Response	450	1203	1255	1231.31	7.71	0.75	3.61
Follow Directions	5,529	1203	1290	1239.65	9.92	0.88	3.35
Special School	300	1203	1250	1232.24	8.37	0.80	3.56
Regular School Self-contained	3,792	1203	1290	1238.35	9.77	0.87	3.37
Regular School Primarily Self-contained	1,167	1216	1283	1240.66	9.60	0.88	3.30
Regular School Resource Room	462	1221	1283	1242.88	11.07	0.89	3.46
Regular School General Education	258	1226	1266	1242.50	9.77	0.88	3.30
Communicates Primarily Through Cries	378	1203	1259	1231.30	7.96	0.77	3.62
Uses Intentional Communication	1,512	1203	1290	1235.59	8.73	0.84	3.37
Uses Symbolic Language	4,089	1203	1290	1241.01	9.95	0.88	3.34

Table M-10. IRT Subgroup Reliability: Mathematics Grade 5

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	5,883	1200	1290	1241.92	10.42	0.86	3.73
Female	1,950	1200	1288	1241.31	10.26	0.86	3.74
Male	3,912	1200	1290	1242.18	10.46	0.87	3.73
Gender Undefined	21	1235	1270	1249.43	13.38	NA	NA
Hispanic or Latino	1,356	1216	1288	1240.94	9.79	0.85	3.67
American Indian or Alaska Native	210	1227	1262	1241.57	8.00	0.79	3.63
Asian	165	1200	1270	1238.45	10.51	0.86	3.77
Black or African American	1,167	1207	1290	1243.60	11.38	0.88	3.81
Native Hawaiian or Pacific Islander	57	1227	1267	1240.84	9.65	NA	NA
White (non-Hispanic)	2,484	1200	1288	1241.57	9.87	0.86	3.68
Two or More Races (non-Hispanic)	261	1200	1279	1241.15	11.93	0.85	4.05
No Primary race/Ethnicity Undefined	1,356	1216	1288	1240.94	9.79	0.85	3.67
Currently receiving LEP services	429	1223	1270	1239.88	8.70	0.83	3.59
Not receiving LEP services	3,474	1200	1290	1242.19	10.53	0.86	3.76
Economically Disadvantaged Students	1,728	1200	1279	1242.68	10.38	0.86	3.76
Non-economically Disadvantaged Students	1,917	1207	1290	1241.40	10.25	0.86	3.72
Migrant	165	1207	1267	1237.51	9.81	0.86	3.62
Non-migrant	3,432	1200	1290	1242.17	10.34	0.86	3.75
Augmentative Communication	1,068	1200	1288	1237.50	9.20	0.83	3.63
No Augmentative Communication	4,767	1200	1290	1242.87	10.36	0.86	3.75
Undefined Augmentative Communications	48	1226	1288	1245.44	15.33	NA	NA
Hearing Loss	120	1216	1264	1240.38	9.21	0.84	3.60
Within Normal Limits	5,763	1200	1290	1241.95	10.44	0.86	3.74
Visual Impairment	225	1207	1270	1237.77	10.97	0.88	3.74
Within Normal Limits	5,619	1200	1290	1242.10	10.37	0.86	3.73
Undefined Visual Impairment	39	1227	1264	1239.77	10.02	NA	NA
Sensory Stimuli Response	393	1200	1264	1235.90	9.85	0.82	3.79
Follow Directions	5,490	1200	1290	1242.35	10.33	0.86	3.73
Special School	303	1212	1279	1237.66	8.81	0.83	3.62
Regular School Self-contained	3,573	1200	1290	1241.16	10.41	0.86	3.74
Regular School Primarily Self-contained	1,251	1221	1288	1243.20	9.81	0.85	3.68
Regular School Resource Room	462	1200	1279	1243.58	10.75	0.87	3.80
Regular School General Education	294	1226	1279	1247.41	10.92	0.87	3.85
Communicates Primarily Through Cries	336	1200	1260	1235.55	9.88	0.81	3.85
Uses Intentional Communication	1,305	1200	1279	1239.26	10.22	0.86	3.67
Uses Symbolic Language	4,242	1207	1290	1243.24	10.18	0.86	3.74

Table M-11. IRT Subgroup Reliability: Mathematics Grade 6

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	6,027	1209	1290	1241.66	11.04	0.89	3.22
Female	1,983	1209	1290	1240.97	10.40	0.89	3.15
Male	4,026	1209	1290	1242.01	11.35	0.90	3.25
Gender Undefined	18	1231	1246	1238.67	5.27	NA	NA
Hispanic or Latino	1,473	1209	1290	1240.82	10.37	0.89	3.13
American Indian or Alaska Native	276	1224	1290	1242.01	11.45	0.86	3.37
Asian	138	1227	1271	1239.61	9.42	0.89	2.95
Black or African American	1,173	1209	1290	1242.80	12.06	0.89	3.40
Native Hawaiian or Pacific Islander	48	1229	1256	1238.50	8.58	NA	NA
White (non-Hispanic)	2,451	1209	1290	1241.51	10.83	0.90	3.18
Two or More Races (non-Hispanic)	282	1226	1290	1240.64	10.09	0.87	3.11
No Primary race/Ethnicity Undefined	1,473	1209	1290	1240.82	10.37	0.89	3.13
Currently receiving LEP services	441	1219	1278	1239.97	9.61	0.89	3.02
Not receiving LEP services	3,381	1209	1290	1242.20	11.39	0.89	3.27
Economically Disadvantaged Students	1,644	1209	1290	1242.86	11.75	0.88	3.38
Non-economically Disadvantaged Students	1,926	1219	1290	1241.45	11.07	0.90	3.18
Migrant	174	1222	1278	1243.47	11.99	0.90	3.40
Non-migrant	3,363	1209	1290	1241.90	11.21	0.90	3.23
Augmentative Communication	999	1209	1266	1235.96	7.69	0.85	2.92
No Augmentative Communication	4,977	1209	1290	1242.81	11.27	0.89	3.28
Undefined Augmentative Communications	51	1224	1260	1240.82	9.58	NA	NA
Hearing Loss	96	1226	1266	1239.56	10.68	NA	NA
Within Normal Limits	5,931	1209	1290	1241.69	11.04	0.89	3.22
Visual Impairment	225	1215	1278	1236.64	10.01	0.89	3.10
Within Normal Limits	5,766	1209	1290	1241.85	11.04	0.89	3.22
Undefined Visual Impairment	36	1226	1263	1241.83	10.62	NA	NA
Sensory Stimuli Response	369	1209	1263	1234.67	8.85	0.86	3.10
Follow Directions	5,658	1209	1290	1242.11	11.02	0.89	3.23
Special School	348	1215	1263	1235.48	6.99	0.82	2.90
Regular School Self-contained	3,942	1209	1290	1240.89	10.96	0.89	3.20
Regular School Primarily Self-contained	1,137	1215	1278	1243.11	10.48	0.89	3.18
Regular School Resource Room	360	1227	1290	1248.15	11.71	0.87	3.65
Regular School General Education	240	1230	1290	1246.56	11.43	0.86	3.56
Communicates Primarily Through Cries	378	1209	1263	1234.45	7.59	0.83	2.99
Uses Intentional Communication	1,212	1209	1290	1237.61	9.55	0.87	3.07
Uses Symbolic Language	4,437	1219	1290	1243.38	11.13	0.89	3.28

Table M-12. IRT Subgroup Reliability: Mathematics Grade 7

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	6,135	1206	1290	1241.70	12.13	0.89	3.75
Female	2,193	1206	1290	1240.49	11.19	0.88	3.66
Male	3,912	1206	1290	1242.37	12.58	0.89	3.81
Gender Undefined	30	1231	1266	1242.50	13.00	NA	NA
Hispanic or Latino	1,500	1206	1290	1240.14	11.60	0.89	3.66
American Indian or Alaska Native	234	1217	1275	1240.60	10.70	0.88	3.59
Asian	111	1217	1284	1240.30	13.13	0.90	3.85
Black or African American	1,209	1211	1290	1243.47	12.94	0.89	3.89
Native Hawaiian or Pacific Islander	57	1224	1275	1238.53	14.09	NA	NA
White (non-Hispanic)	2,589	1206	1290	1241.93	11.75	0.89	3.73
Two or More Races (non-Hispanic)	252	1219	1266	1238.77	10.10	0.88	3.54
No Primary race/Ethnicity Undefined	1,500	1206	1290	1240.14	11.60	0.89	3.66
Currently receiving LEP services	324	1214	1290	1239.70	11.30	0.88	3.65
Not receiving LEP services	3,585	1206	1290	1242.44	12.22	0.89	3.80
Economically Disadvantaged Students	1,722	1206	1290	1243.10	12.33	0.88	3.85
Non-economically Disadvantaged Students	2,025	1206	1290	1241.64	11.82	0.89	3.73
Migrant	198	1211	1290	1240.12	14.92	0.89	4.17
Non-migrant	3,507	1206	1290	1242.40	11.91	0.88	3.77
Augmentative Communication	912	1211	1266	1233.88	8.78	0.83	3.55
No Augmentative Communication	5,187	1206	1290	1243.03	12.11	0.89	3.79
Undefined Augmentative Communications	36	1233	1284	1247.33	12.90	NA	NA
Hearing Loss	126	1214	1275	1236.67	11.36	0.89	3.69
Within Normal Limits	6,009	1206	1290	1241.80	12.12	0.89	3.75
Visual Impairment	225	1217	1257	1233.96	8.64	0.83	3.49
Within Normal Limits	5,883	1206	1290	1241.97	12.12	0.89	3.76
Undefined Visual Impairment	27	1219	1270	1246.22	16.30	NA	NA
Sensory Stimuli Response	339	1211	1261	1231.78	8.87	0.84	3.57
Follow Directions	5,796	1206	1290	1242.28	12.04	0.89	3.76
Special School	327	1211	1261	1234.29	9.17	0.85	3.51
Regular School Self-contained	4,086	1206	1290	1240.76	11.67	0.89	3.68
Regular School Primarily Self-contained	1,110	1206	1290	1243.68	12.34	0.87	3.91
Regular School Resource Room	381	1224	1290	1247.33	12.63	0.87	4.04
Regular School General Education	228	1227	1290	1250.00	12.50	0.84	4.22
Communicates Primarily Through Cries	312	1211	1275	1231.66	9.30	0.84	3.63
Uses Intentional Communication	1,221	1211	1284	1236.84	10.50	0.88	3.59
Uses Symbolic Language	4,602	1206	1290	1243.67	12.01	0.88	3.81

Table M-13. IRT Subgroup Reliability: Mathematics Grade 8

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	6,255	1201	1290	1241.78	10.85	0.85	3.70
Female	2,349	1201	1290	1241.01	10.14	0.85	3.59
Male	3,876	1201	1290	1242.26	11.24	0.85	3.77
Gender Undefined	30	1228	1262	1240.90	9.58	NA	NA
Hispanic or Latino	1,365	1201	1290	1241.13	10.90	0.84	3.71
American Indian or Alaska Native	240	1222	1271	1240.95	9.82	0.86	3.50
Asian	126	1220	1271	1240.69	9.98	0.87	3.55
Black or African American	1,215	1220	1290	1243.18	11.06	0.84	3.78
Native Hawaiian or Pacific Islander	57	1201	1255	1234.89	12.18	NA	NA
White (non-Hispanic)	2,880	1214	1290	1241.48	10.71	0.85	3.67
Two or More Races (non-Hispanic)	213	1222	1290	1242.34	10.10	0.81	3.72
No Primary race/Ethnicity Undefined	1,365	1201	1290	1241.13	10.90	0.84	3.71
Currently receiving LEP services	327	1220	1290	1241.13	11.03	0.85	3.68
Not receiving LEP services	3,837	1201	1290	1241.75	10.78	0.85	3.70
Economically Disadvantaged Students	1,806	1214	1290	1242.30	10.54	0.84	3.69
Non-economically Disadvantaged Students	2,184	1217	1290	1241.36	10.67	0.85	3.68
Migrant	210	1224	1290	1243.60	12.76	0.82	4.18
Non-migrant	3,726	1214	1290	1241.66	10.49	0.85	3.66
Augmentative Communication	861	1214	1266	1235.77	7.94	0.82	3.37
No Augmentative Communication	5,331	1201	1290	1242.77	10.97	0.85	3.76
Undefined Augmentative Communications	63	1227	1262	1240.43	8.27	NA	NA
Hearing Loss	138	1224	1259	1239.00	8.85	0.85	3.39
Within Normal Limits	6,117	1201	1290	1241.85	10.88	0.85	3.71
Visual Impairment	204	1217	1290	1237.28	12.39	0.87	3.86
Within Normal Limits	6,027	1201	1290	1241.94	10.76	0.85	3.70
Undefined Visual Impairment	24	1224	1255	1240.75	9.63	NA	NA
Sensory Stimuli Response	318	1201	1259	1234.06	8.50	0.82	3.47
Follow Directions	5,937	1201	1290	1242.20	10.80	0.85	3.71
Special School	351	1220	1290	1236.03	8.66	0.79	3.51
Regular School Self-contained	4,014	1201	1290	1240.75	10.23	0.85	3.61
Regular School Primarily Self-contained	1,254	1201	1290	1244.43	11.11	0.85	3.82
Regular School Resource Room	420	1220	1290	1246.98	13.48	0.83	4.41
Regular School General Education	216	1226	1279	1244.96	9.83	0.84	3.67
Communicates Primarily Through Cries	258	1201	1249	1232.94	8.76	0.82	3.55
Uses Intentional Communication	1,188	1214	1290	1237.34	10.13	0.85	3.58
Uses Symbolic Language	4,809	1220	1290	1243.36	10.59	0.84	3.74

Table M-14. IRT Subgroup Reliability: Mathematics Grade HS

Description	Number of Students	Scale Score Minimum	Scale Score Maximum	Scale Score Mean	Scale Score Standard Deviation	IRT Marginal Reliability	Standard Error
All	5,826	1200	1290	1242.07	8.95	0.84	3.07
Female	2,034	1210	1290	1242.15	8.41	0.83	3.02
Male	3,675	1200	1290	1241.98	9.20	0.85	3.10
Gender Undefined	117	1231	1277	1243.74	10.17	0.85	3.30
Hispanic or Latino	1,185	1219	1290	1242.06	8.50	0.85	3.00
American Indian or Alaska Native	186	1210	1277	1242.52	10.29	0.88	3.23
Asian	111	1219	1260	1240.95	8.90	0.89	2.90
Black or African American	1,077	1200	1290	1242.56	9.37	0.81	3.21
Native Hawaiian or Pacific Islander	27	1233	1255	1242.44	7.27	NA	NA
White (non-Hispanic)	2,784	1216	1290	1241.67	8.48	0.84	3.00
Two or More Races (non-Hispanic)	216	1210	1290	1241.00	10.60	0.84	3.30
No Primary race/Ethnicity Undefined	1,185	1219	1290	1242.06	8.50	0.85	3.00
Currently receiving LEP services	219	1216	1255	1240.40	6.66	0.82	2.80
Not receiving LEP services	3,612	1200	1290	1242.37	8.83	0.83	3.09
Economically Disadvantaged Students	1,524	1210	1277	1242.66	8.25	0.85	3.00
Non-economically Disadvantaged Students	2,229	1200	1290	1242.09	9.14	0.81	3.14
Migrant	117	1210	1255	1237.28	8.10	0.86	2.92
Non-migrant	3,624	1200	1290	1242.48	8.75	0.82	3.09
Augmentative Communication	705	1200	1277	1237.89	8.59	0.81	3.02
No Augmentative Communication	5,088	1210	1290	1242.63	8.86	0.84	3.08
Undefined Augmentative Communications	33	1236	1260	1245.82	6.89	NA	NA
Hearing Loss	132	1216	1264	1239.05	9.27	0.89	2.98
Within Normal Limits	5,694	1200	1290	1242.14	8.93	0.84	3.08
Visual Impairment	165	1200	1277	1240.35	12.23	0.78	3.76
Within Normal Limits	5,646	1210	1290	1242.14	8.84	0.84	3.05
Undefined Visual Impairment	15	1228	1244	1237.80	5.57	NA	NA
Sensory Stimuli Response	237	1200	1255	1234.43	9.30	0.72	3.40
Follow Directions	5,589	1210	1290	1242.40	8.79	0.84	3.06
Special School	417	1210	1277	1237.60	9.05	0.87	3.00
Regular School Self-contained	3,663	1200	1290	1241.92	9.02	0.83	3.10
Regular School Primarily Self-contained	1,191	1216	1277	1242.23	7.95	0.85	2.95
Regular School Resource Room	432	1232	1277	1246.05	8.45	0.83	3.22
Regular School General Education	123	1230	1277	1246.46	9.47	0.85	3.33
Communicates Primarily Through Cries	243	1200	1255	1235.16	9.52	0.74	3.40
Uses Intentional Communication	1,014	1210	1277	1238.77	7.78	0.86	2.86
Uses Symbolic Language	4,569	1216	1290	1243.18	8.82	0.83	3.10

APPENDIX N
DECISION ACCURACY AND CONSISTENCY RESULTS

Table N-1. Summary of Decision Accuracy (and Consistency) Results by Content Area and Grade—Overall and Conditional on Performance Level

				Conditional on Level			
Content Area	Grade	Overall	Kappa	Level 1	Level 2	Level 3	Level 4
ELA	3	0.83 (0.76)	0.66	0.89 (0.87)	0.65 (0.48)	0.84 (0.80)	0.81 (0.74)
	4	0.83 (0.77)	0.67	0.90 (0.87)	0.68 (0.55)	0.85 (0.80)	0.79 (0.73)
	5	0.82 (0.75)	0.66	0.88 (0.83)	0.76 (0.67)	0.83 (0.76)	0.82 (0.73)
	6	0.80 (0.72)	0.60	0.85 (0.78)	0.71 (0.59)	0.83 (0.78)	0.82 (0.71)
	7	0.81 (0.74)	0.61	0.92 (0.83)	0.57 (0.46)	0.83 (0.80)	0.83 (0.69)
	8	0.80 (0.72)	0.61	0.86 (0.80)	0.76 (0.66)	0.79 (0.71)	0.81 (0.75)
	HS	0.81 (0.73)	0.61	0.85 (0.78)	0.59 (0.47)	0.85 (0.80)	0.85 (0.77)
Mathematics	3	0.78 (0.69)	0.56	0.84 (0.81)	0.61 (0.44)	0.79 (0.74)	0.82 (0.68)
	4	0.74 (0.65)	0.51	0.75 (0.63)	0.66 (0.59)	0.80 (0.68)	0.84 (0.77)
	5	0.76 (0.67)	0.52	0.81 (0.66)	0.66 (0.56)	0.78 (0.71)	0.89 (0.77)
	6	0.75 (0.66)	0.54	0.77 (0.61)	0.65 (0.58)	0.78 (0.67)	0.90 (0.83)
	7	0.76 (0.68)	0.56	0.84 (0.75)	0.58 (0.47)	0.77 (0.71)	0.88 (0.80)
	8	0.77 (0.68)	0.55	0.84 (0.77)	0.56 (0.42)	0.77 (0.72)	0.86 (0.76)
	HS	0.78 (0.69)	0.55	0.84 (0.73)	0.57 (0.44)	0.81 (0.76)	0.86 (0.75)

Table N-2. Summary of Decision Accuracy (and Consistency) Results by Content Area and Grade—Conditional on Cutpoint

Content Area	Grade	Level 1/Level 2				Level 2/Level 3				Level 3/Level 4			
		Accuracy	Consistency	False Positive	False Negative	Accuracy	Consistency	False Positive	False Negative	Accuracy	Consistency	False Positive	False Negative
ELA	3	0.94	0.91	0.04	0.02	0.93	0.91	0.03	0.04	0.96	0.94	0.02	0.02
	4	0.94	0.91	0.03	0.03	0.93	0.90	0.03	0.04	0.96	0.95	0.02	0.02
	5	0.93	0.90	0.03	0.03	0.93	0.90	0.03	0.03	0.96	0.94	0.02	0.02
	6	0.94	0.92	0.03	0.03	0.91	0.88	0.04	0.05	0.95	0.93	0.03	0.02
	7	0.94	0.91	0.02	0.04	0.92	0.89	0.03	0.05	0.94	0.92	0.04	0.02
	8	0.94	0.91	0.03	0.03	0.91	0.88	0.04	0.04	0.95	0.93	0.03	0.03
	HS	0.93	0.90	0.03	0.04	0.92	0.88	0.04	0.04	0.95	0.93	0.03	0.02
Mathematics	3	0.93	0.91	0.04	0.02	0.90	0.86	0.04	0.06	0.94	0.92	0.03	0.02
	4	0.89	0.84	0.05	0.06	0.90	0.85	0.07	0.03	0.96	0.94	0.02	0.02
	5	0.93	0.90	0.03	0.05	0.88	0.83	0.06	0.06	0.95	0.93	0.03	0.01
	6	0.89	0.85	0.03	0.07	0.90	0.87	0.06	0.04	0.95	0.94	0.03	0.02
	7	0.90	0.87	0.04	0.06	0.90	0.86	0.05	0.05	0.95	0.93	0.03	0.02
	8	0.93	0.90	0.04	0.04	0.88	0.84	0.04	0.07	0.95	0.92	0.03	0.02
	HS	0.93	0.90	0.02	0.04	0.89	0.85	0.05	0.06	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 O
LIST OF ACRONYMS AND COMMONLY USED TERMS

Acronyms Used in the 2023 MSAA Technical Report	
2PL	two-parameter logistic
AA-AAS	Alternate Assessment Aligned with Alternate Achievement Standards (utilized under ESEA until 2015)
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
APIP	Accessible Portable Item Protocol
BIE	Bureau of Indian Education
CBT	Computer-based test
CCC	Core Content Connector
CCSS	Common Core State Standards
CSEM	conditional standard error of measurement
CTT	Classical Test Theory
DAC	decision accuracy and consistency
DETECT	Dimensionality Evaluation to Enumerate Contributing Traits
DIF	differential Item functioning
DIMTEST	computer program used by Cognia
DNU	do not use
DTA	Directions for Test Administration
ELA	English language arts
EOTS	end-of-test survey
ESEA	Elementary and Secondary Education Act
ESR	early stopping rule
ESSA	Every Student Succeeds Act
EU	essential understanding
FKSA	focal knowledge, skills, and ability
GM	geometry (in standards)
GRM	graded-response model
HOSS	highest obtainable scale score
ICC	item characteristic curve
ICCC	item category characteristic curve
ICTC	item category threshold curve
IEP	individualized education program
IIF	Item information function
IRC	Item Review Committee (includes Content Review Committee and Bias-Sensitivity Review Committee)
IRT	Item Response Theory
IT	information technology
KSA	knowledge, skills, and ability

continued

Acronyms Used in the 2023 MSAA Technical Report	
LEP	limited English proficiency
LOSS	lowest obtainable scale score
LPF	Learning Progression Framework
MSAA	Multi-State Alternate Assessment
NCME	National Council on Measurement in Education
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
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
SD	standard deviation
SEM	standard error of measurement
SIU	score interpretations and uses
SQA	Software Quality Assurance
SRC	student response check
STL	scoring team leader
TA	test administrators
TAC	Technical Advisory Committee
TAM	Test Administration Manual
TC	test coordinators
TCC	test characteristic curve
TIF	test information function
UWC	use with caution
WRCC	Writing Common—a code within iScore that is used to identify items for scoring
NCSC	National Center and State Collaborative

Commonly Used Terms	
a	The item response theory index of item discrimination, analogous to the point-biserial and biserial correlations in classical test theory. It reflects the slope of the item response function. Often ranging from 0.1 to 2.0 in practice, a higher value indicates a better-performing item.
Accreditation	Accreditation by an outside agency affirms that an organization has met a certain level of standards. Certification testing programs may become accredited by meeting specified standards in test development, psychometrics, bylaws, management, etc.
b	The item response theory index of item difficulty or location, analogous to the P-value (P+) of classical test theory. Typically ranging from -3.0 to 3.0 in practice, a higher value indicates a more difficult item.
Biserial Correlation	A classical index of item discrimination, highly similar to the more commonly used point-biserial. The biserial correlation assumes that the item scores and test scores reflect an underlying normal distribution, which is not always the case.
Blueprint	A test blueprint, or test specification, details how an exam is to be constructed. It includes important information, such as the total number of items, the number of items in each content area or domain, the number of items that are recall vs. reasoning, and the item formats to be utilized.
c	The item response theory pseudo-guessing parameter, representing the lower asymptote of the item response function. It is theoretically near the value of 1/k, where k is the number of alternatives. For example, with the typical four-option multiple choice item, a candidate has a base chance of 25% of guessing the correct answer.
Certification	A non-mandatory testing program which certifies that candidates have achieved a minimum standard or knowledge or performance.
Classical Test Theory	A psychometric analysis and test development paradigm based on correlations, proportions, and other statistics that are relatively simple compared to IRT. It is therefore more appropriate for smaller samples, especially for fewer than 100.
Classification	The use of tests for classifying candidates into categories, such as pass/fail, non-master/master, or basic/proficient/advanced.
Computerized Adaptive Testing	A dynamic method of test administration where items are selected one at a time to match item difficulty and candidate ability as closely as possible. This helps prevent candidates being presented with items that are too difficult or too easy for them, which has multiple benefits. Often, the test only takes half as many items to obtain a similar level of accuracy to form-based tests. This reduces the testing time per examinee and reduces the total number of times an item is exposed, as well as increasing security by the fact that nearly every candidate will receive a different set of items.
Cutscore	Also known as a passing score, the cutscore is the score that a candidate must achieve to obtain a certain classification, such as “pass” on a licensure or certification exam.
Criterion-Referenced	A test score (not a test) is criterion-referenced if it is interpreted regarding a specified criterion and not compared to scores of other candidates. For instance, providing the number-correct score does not relate any information regarding a candidate’s relative standing.
Distractors	Distractors are the incorrect options of a multiple-choice item. A distractor analysis is an important part of psychometric review, as it helps determine if one is acting as a keyed response.
Equating	The process of determining comparable scores on different forms of an examination. For example, if Form A is more difficult than Form B, it might be desirable to adjust scores on Form A upward for the purposes of comparing them to scores on Form B. Usually, this is done statistically based on items that are on both forms, which are called equater, anchor, or common items. Since the groups who took the two forms are different, this is called a common items non-equivalent groups design.

continued

Commonly Used Terms	
Form	A specific set of items that are administered together for a test. For example, if a test included a certain set of 100 items this year, and a different set of 100 items next year, these would be two distinct forms.
Item	The basic component of a test, often colloquially referred to as a "question," but items are not necessarily phrased as a question. They can be as varied as true/false statements, rating scales, and performance task simulations, in addition to the ubiquitous multiple-choice item.
Item Bank	A repository of items for a testing program, including items at all stages, such as newly written, reviewed, pretested, active, and retired.
Item Difficulty	A statistical index of how easy/hard the item is with respect to the underlying ability/trait. That is, an item is difficult if not many people get it correct or respond in the keyed direction.
Item Discrimination	A statistical index of the quality of the item, assessing how well it differentiates examinees of high vs. low ability. Items with low discrimination are considered poor quality and are candidates to be revised or retired.
Item Response Theory (IRT)	A comprehensive approach to psychometric analysis and test development that utilizes complex mathematical models. This provides several benefits, including the ability to design CATs, but requires larger sample sizes. A common rule of thumb is 100 candidates for the one-parameter model and 500 for the three-parameter model.
Key	The key is the correct response to an item.
Knowledge, Skills, and Abilities (KSAs)	A critical step in testing for either employment or professional credentials is to determine the KSAs that are important in a job. This is often done via a job analysis study.
Licensure	A testing program mandated by a government body. The test must be passed to perform the task in question, whether it is to work in the profession or drive a car.
Norm-Referenced	A test score (not a test) is norm-referenced if it is interpreted with regard to the performance of other candidates. Percentile rank is an example of this, because it does not provide any information regarding how many items the candidate got correct.
P-value	A classical index of item difficulty, presented as the proportion of candidates who correctly responded to the item. A value above 0.90 indicates an easy item, while a value below 0.50 indicates a relatively difficult item. Note that it is inverted; a higher value indicates less difficulty.
Point-Biserial Correlation	A classical index of item discrimination, calculated as the Pearson correlation between the item score and the total test score. If below 0.0, low-scoring candidates are doing better than high-scoring candidates, and the item should be revised or retired. Low positive values are marginal, higher positive values are ideal.
Field-test Item	An item that is administered to candidates simply for the purpose of obtaining data for future psychometric analysis. The results on this item are not included in the score. It is often prudent to include a small number of pretest items in a test.
Reliability	A measure of the repeatability or consistency of the measurement process. Often, this is indexed by a single number, most commonly the internal consistency index coefficient alpha or its dichotomous formulation, KR-20. Under most conditions, these range from 0.0 to 1.0, with 1.0 being a perfectly reliable measurement. However, just because a test is reliable does not mean that it is valid, i.e., measures what it is supposed to measure.

continued

Commonly Used Terms	
Scaling	A process of converting scores obtained on an exam to an arbitrary scale. This is done so that all the forms and exams used by a testing organization are on a common scale. For example, suppose an organization had two testing programs, one with 50 items and one with 150 items. All scores could be put on the same scale to standardize score reporting.
Standard-Setting	A formal study conducted by a testing organization to determine standards for a testing program, which are manifested as a cutscore. Common methods include the Angoff, Bookmark, Contrasting Groups, and Borderline Survey methods.
Subject Matter Expert	An extremely vital person in the test development process. SMEs are necessary to write items, review items, participate in standard-setting studies and job analyses, and oversee the testing program to ensure its fidelity to its true intent.
Validity	The concept that test scores can be interpreted as intended. For example, a test for certification in a profession should reflect basic knowledge of that profession, and not intelligence or other constructs, and scores can therefore be interpreted as evidencing professional competence. Validity must be formally established and maintained by empirical studies as well as sound psychometric and test development practices.

APPENDIX P
ELEMENT LEVEL RATING SCALE

Element Level Rating Scale (from Chapter 11)

Assumption 1.1. The content of the test represents the content of the standards (i.e., the Core Content Connectors).

The evidence to support this test alignment assumption and its elements was generated in a series of alignment studies that were conducted between 2012 and 2015 by the National Center and State Collaborative (NCSC). Details regarding the alignment reports and evidence of findings is available in the *National Center and State Collaborative 2015 Operational Assessment Technical Manual* (see http://ncscpartners.org/Media/Default/PDFs/Resources/NCSC15_NCSC_TechnicalManualNarrative.pdf).

Element 1.1.1. The Core Content Connectors are aligned to the states' grade-level academic content standards. (See alignment question 1 in the technical manual.)

> ***Evidence:*** The technical manual summarizes evidence from the Reading, Writing, and Mathematics Relationship Studies and states that "all the evidence suggested the mathematics, reading, and writing CCCs had a strong relationship to the CCSS standards" (p. 75). The evidence in all three content areas includes content centrality and performance centrality, that the overall cognitive complexity of the assessment is appropriately lower than that of the grade-level content standards, and that some Core Content Connectors were rated at high depth of knowledge levels, suggesting that MSAA accesses challenging academic standards.

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 CCCs and each partner's academic content standards may be warranted if the partners' standards are modified.

Element 1.1.2. The 2023 MSAA items are aligned to the Core Content Connectors. (See alignment question 3 in the technical manual.)

> ***Evidence:*** The NCSC technical manual indicates that "NCSC designed the operational items to assess the knowledge and skills of a wide variety of students with the most significant cognitive disabilities [and]…The study provided evidence that the assessment's operational items allowed students using various communication modes and with specific characteristics to access the items. Panelists indicated that the items were suitable for students who used various communication modes, and panelists indicated that no modifications were necessary to enable student access to the test items" (p. 80). Item specifications and development and review processes every year since 2015 are designed and implemented to ensure that items are closely aligned to the Core Content Connectors.

Relevance: Evidence is highly relevant
Completeness: Evidence is complete
Overall Support: Existing evidence strongly supports the element.
Possible Challenges: As various communication technologies evolve; it may be necessary to revisit the suitableness of the items to various communication modes.

Element 1.1.3. The MSAA partners have confirmed that the MSAA Core Content Connectors, which are assessed on the MSAA, 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 CCCs and each partner's academic content standards may be warranted if the partners' standards are modified.

Element 1.1.4. The 2023 operational MSAA items are aligned to the MSAA performance level descriptors. (See alignment question 4 in the technical manual.)

> ***Evidence:*** As reported in the technical manual, "Panelists rated item groups based on their judgment of whether the KSAs in the item groups represent the KSAs in the PLDs… For all content areas and grade levels, the majority of the item groups were rated as having the same KSAs as the PLDs, ranging from 57% to 78%. Some panelists indicated that some KSAs were missing in the item groups or the PLDs, but overall, the overlap of KSAs found in the item groups and PLDs was acceptable" (p. 81). The items are developed following item specifications for each CCC. The item specifications accomplish two purposes: (1) they provide both general and specific guidelines for developing all test items at the grade levels assessed and (2) they describe the test items and prompt types needed. Within the specifications documents are sections dedicated to information about 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, in addition to the items to the performance level descriptors, there is alignment from the item specifications to the performance level descriptors.

Relevance: Evidence is highly relevant
Completeness: Evidence is moderate to substantial
Overall Support: Existing evidence moderately supports the element.
Possible Challenges: A follow-up study of alignment between more recent items and the MSAA performance level descriptors may be warranted.

Assumption 1.2. MSAA 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. Item scoring rubrics and criteria focus on construct-relevant aspects of student response.
Element 1.2.5. Scaffolding is not a source of construct-irrelevant variance.
Element 1.2.6. Item rendering does not interfere with students' access to test content.
Element 1.2.7. Platform does not interfere with students' ability to interact with test content.
Element 1.2.8. Items are free of bias and sensitive issues.

The evidence for Elements 1.2.1 through 1.2.8 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 for each individual element. 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 items are aligned to the Core Content Connectors, through the targeted focal knowledge, skills, and abilities (FKSAs) and/or essential understandings (EUs) that support 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 Core Content Connectors]), the evidence for 1.1.2 is Complete Evidence.

Evidence for 1.2.1, 1.2.2, 1.2.3, 1.2.4, 1.2.5, and 1.2.8: During the item development process, the items followed a rigorous development cycle, including reviews by MSAA State Representatives and by Item Content and Bias and Sensitivity panelists. See Chapter 3 for a detailed description of the item review process.

Evidence for 1.2.1, 1.2.2, 1.2.3, 1.2.4, 1.2.6, and 1.2.7: Cognitive labs (also referred to by NCSC as Student Interaction Studies) were conducted by NCSC in the early stages of development of the assessment program to acquire detailed information about the cognitive processes used by students in responding to assessment tasks. The studies resulted in adjustments in the assessment program to ensure the construct validity of student response data. This information is outlined in the *National Center and State Collaborative 2015 Operational Assessment Technical Manual*.

Evidence for 1.2.5, 1.2.6, and 1.2.7: Three usability studies were conducted by NCSC in the early stages of development of the assessment program to evaluate how students and teachers interacted with items and gathered evidence related to item complexity and usability. The studies resulted in adjustments to ensure the assessment met all usability standards required to support the validity of the assessment program. This information is outlined in the *National Center and State Collaborative 2015 Operational Assessment Technical Manual*.

Evidence for 1.2.1, 1.2.2, 1.2.3, 1.2.4, 1.2.5, 1.2.6 and 1.2.8: During the item development process, the items followed a rigorous development cycle, including reviews by MSAA State Representatives and by Item Content and Bias and Sensitivity panelists. See Chapter 3 for a detailed description of the item review process.

Evidence for 1.2.1, 1.2.2, 1.2.3, 1.2.4, 1.2.6, and 1.2.7: Cognitive labs (also referred to by NCSC as Student Interaction Studies) were conducted by NCSC in the early stages of development of the assessment program to acquire detailed information about the cognitive processes used by students in responding to assessment tasks. The studies resulted in adjustments in the assessment program to ensure the construct validity of student response data. This information is outlined in the *National Center and State Collaborative 2015 Operational Assessment Technical Manual*.

Evidence for 1.2.8: 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 H.

Summary of evidence for 1.2.1:

Relevance: The evidence is highly relevant
Completeness: The evidence is *Moderate to Substantial*
Overall Support: The existing evidence strongly supports the element
Possible Challenges: If the standard setting included a process in which subject matter experts evaluated the KSA demands of the items relative to the KSAs in the PLDs, that would provide additional evidence.

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 element
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 element
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 element
Possible Challenges: A description of the standard setting process where subject matter experts evaluated the KSA demands of the scoring rubrics and criteria relative to the KSAs in the PLDs would provide additional evidence.

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 element
Possible Challenges: A follow-up study evaluating whether construct-irrelevant variance of more recent items may be warranted.

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 element
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 1.2.7:

Relevance: The evidence is highly relevant
Completeness: The evidence is *Moderate to Substantial*
Overall Support: The existing evidence strongly supports the element
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 1.2.8:

Relevance: The evidence is highly relevant
Completeness: The evidence is moderate to substantial
Overall Support: The existing evidence strongly supports the element
Possible Challenges: Results from the NCSC Student Interaction studies may contain evidence of bias or sensitivity issues reported during that study.

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 participate in mandatory test administration training each year, including 2023. 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 are required to use: the Test Administrator Manual (TAM), the Directions for Test Administration (DTA), and the MSAA Online Assessment System User Guide for Test Administrators. 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 are 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:

- receives four best-practice videos;
- receives a technical support chart that provides examples of when and who to contact to obtain answers concerning MSAA assessment or administration; and
- completes 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 2023 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 Arizona Department of Education summarized results from its spring 2022 MSAA administration observations. Their observations included the following:

- Of 126 responses, 92.9% administered the MSAA following the instructions in the Directions for Test Administration (DTA).
- Of 118 responses, 94.9% observed secure storage of secure test materials.

Relevance: The evidence is highly relevant
Completeness: The evidence is limited

Overall Support: The existing evidence strongly supports the element
Possible Challenges: Additional information on results from observations of 2023 test administrations for more MSAA partners would bolster this claim.

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:

o **ELA and Mathematics**: AZ had one instance of an incident report. The report did not indicate a need for state investigation and had minimal impact on testing for one student.

Relevance: Highly relevant
Completeness: Limited Evidence.
Overall Support: Evidence provides limited support of the claim.
Possible Challenges: Not all states and entities have submitted incident reports.

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

Element 1.4.1. MSAA scores and categorizations into 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.
- 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 described 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 N.

Relevance: Evidence is highly relevant
Completeness: Evidence is complete
Overall Support: Existing evidence strongly supports the element.
Possible Challenges: The impacts on the reliability evidence that could stem from the local item dependence detected in the dimensionality analysis. 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 the 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; Yen, 1993).

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. For polytomously scored items, namely the writing traits, the item parameter estimates include the discrimination parameter, the overall difficulty level parameter, and the step difficulty parameters for each of the possible non-zero scores. 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 psychometric subcommittee. 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 are reviewed for their item information function (IIF) contributions at the performance level cuts to evaluate and rate the quality of each item. After each operational administration, dimensionality analyses are also conducted to determine 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 for paths A, B, and C support intended interpretations about all students who take the MSAA.

Evidence: Evidence for 1.4.3 includes the following:

- Dimensionality: Dimensionality analysis was conducted on each path for 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. Mostly small to moderate violations of local independence were noted, and interpretations of these results were presented. 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 path.
- 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 paths for all the MSAA tests. 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 in ELA and mathematics, psychometric information is highest around the performance levels 2 and 3 cut scores for paths A and B. 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 2 and 3 cut scores for paths A and B is by design. The TIFs for the performance level 4 cut scores in path C tend to be highest below the level 4 cut score, rather than around the level 4 cut score. This reflects the limited number of high difficulty items currently in the MSAA item bank. The amount of psychometric information around these cut scores and about the students whose test performance locates them in that same area is adequate (rule of thumb: information should be greater than or equal to 10) but not maximized. Cognia is working on developing larger numbers of difficult items.

Relevance: Evidence is highly relevant.
Completeness: Evidence completeness is moderate to substantial.
Overall Support: Moderately supports the claim. The TIFs indicate that path-specific information functions are maximized at three different locations on the theta scale (i.e., the cut scores). Another piece of evidence that enhances this argument is the overlap of path-specific TIFs displayed in Appendix J. Still another piece of evidence is the small amount of overlap in the path-specific student proficiency distributions displayed in Tables 9-10 and 9-11. Finally, the path-specific TCC locations are ordered logically and are reasonably separated, as shown in Appendix J. Taken together, all this evidence provides a nearly complete argument supporting this element.
Possible Challenges: Cut 3 TIF is generally adequate but also generally much less than for Cuts 1 and 2. The Cut 3 TIF values have tended to increase to the benchmark of 10.0, but some tests have currently asymptoted to that value while attempts continue to work on increasing Cut 3 TIF for those tests.

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 H.
- Dimensionality: The scale used for reporting scores is a unidimensional scale. Dimensionality analysis was conducted on each Path 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: The few 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 ICCs 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: Evidence to support 1.4.5 includes the following:

- Evaluation of equating items: The appropriateness of the equating is contingent upon the substantive and statistical quality of the equating items. Section 9.2 gives a detailed description of the procedures used to ensure the quality of the equating items, in terms of both content representativeness, as well as statistical stability.
- Third-party analysis: As a check on the equating procedures, a third-party vendor is contracted to conduct the equating analysis. The third-party results are not included in the technical report. A separate report has been written up by the third party and is available upon request.

Relevance: Highly relevant.
Completeness: Evidence is moderately complete
Overall Support: Moderate to substantial support.
Possible Challenges: Perhaps more results could be presented, if it was deemed a high enough priority compared to other elements.

Element 1.4.6. Stage 1 covers a broad enough range of item difficulty and item cognitive complexity to route students into appropriate stage 2 test levels.
Element 1.4.7. Routing into the stage 2 test level is appropriate for students.
Element 1.4.8. Stage 2 test levels are sufficiently separable and targeted toward different ranges of achievement for the MSAA students who are routed to those levels.

Evidence: Evidence for 1.4.6, 1.4.7, and 1.4.8 includes the following.

- Test construction process: The evidence most pertinent to the stages is the report *MSAA Test Construction Process for 2023* (which also reflects the process and criteria for the 2018 MSAA). This detailed report describes how item and test information is targeted for the various stages, including the determination of the routing rules. A brief description is given in Chapter 4. Refer to the full report for a more detailed description.
- Performance level distributions by test path: The test-level statistical results in the technical report are focused on the test as a whole. Thus, the statistical results are focused on Paths A, B, and C, rather than on the stages. Still, some of the path results are especially pertinent to evaluating the psychometric characteristics of the stages. In particular, Section 9.6 reports the Performance Level Distributions across the different paths. If the stages are properly constructed and the routing is properly implemented, the performance level distributions should differ across paths in reasonable ways. The results presented in Section 9.6 can be evaluated in this regard.

Relevance: The evidence is highly relevant
Completeness: Complete evidence
Overall Support: Evidence strongly supports the claim
Possible Challenges: None

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

Machine-scored items are verified for accuracy through a key validation process, while constructed-response item scoring adheres to industry standards with double-blind scoring and comprehensive procedures, all strongly supported by relevant and complete evidence without challenges.

Element 1.5.1: Machine-scored items are verified for accuracy through a key validation process, detailed in Chapter 6. This ensures that designated key responses for operational multiple-choice items are correct. The evidence is highly relevant, complete, and strongly supportive, with no challenges identified.

Element 1.5.2: Constructed-response item scoring meets industry standards, employing rigorous practices like double-blind scoring. All student responses are independently evaluated by two scorers, with about 5.5% undergoing a quality check by the Scoring Team Leader. Chapter 6.2 extensively outlines the procedures, including training, benchmarking, scorer qualifications, leadership, specific scoring rules, quality control, reports, and interrater reliability. The evidence is highly relevant, complete, and strongly supportive, with no challenges noted.

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

Element 1.5.2. Constructed-response item scoring training and monitoring procedures met industry standards.

> ***Evidence:*** Scorer recruitment, training, and qualification and scoring monitoring procedures follow industry standards. There is 100% double blind scoring for the MSAA, therefore all student responses are read independently by two different scorers. Approximately 5.5% of the student responses will also have a read-behind, which served as a quality control check completed by the Scoring Team Leader (STL) on a response previously read by a scorer. This read-behind is also conducted under "blind" scoring parameters. Section 6.2, Open-Response Writing Prompts Scoring Processes, is predominantly devoted to describing all the procedures that are used to ensure the accuracy of the scoring for the constructed-response items, including administrator training and monitoring (6.2.1), benchmarking and identification of scoring materials (6.2.2), scorer recruitment and qualifications (6.2.3), scoring leadership (6.2.4), qualification (6.2.6), specific scoring rules to ensure accuracy (6.2.5), monitoring of quality control (6.2.8), quality reports (6.2.9), and interrater reliability (6.2.10).

Relevance: The evidence is highly relevant
Completeness: Complete evidence
Overall Support: Evidence strongly supports the claim
Possible Challenges: None

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

Element 1.6.1. MSAA 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 disattenuated correlations between 2023 MSAA ELA and mathematics scale scores in grades 3-8 and HS are, in order, .87, .84, .85., .88, .85, .86, and .88. The strong positive values of the MSAA ELA and mathematics correlations provide convergent validity evidence in the sense that they suggest that students' general academic and communicative capabilities are reflected strongly in both their 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: Given the challenge of acquiring additional evidence of student achievement beyond MSAA, no plans are in place to develop additional evidence.

PRIMARY INTENDED SCORE USES

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

Element 2.1.1. MSAA scale scores for groups of students 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 score interpretations was presented above under Assumptions 1.1–1.6 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 M 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.

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: The evidence is highly relevant
Completeness: Moderate to substantial evidence
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.

Element 2.1.2. MSAA 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 N.
- Performance level distributions by testing path: Section 9.6 reports the Performance Level Distributions across the different paths. If the stages are properly constructed and the routing is properly implemented, the performance level distributions should differ across paths in reasonable ways. The results presented in Section 9.6 can be evaluated in this regard.
- For example, in interpreting Tables 9-10 and 9-11, we note that the percentages of examinees classified in Levels 3 and 4 increase across Paths A, B, and C. This reflects the increasing difficulty of the stage-two test forms across Paths A, B, and C.

Relevance: The evidence is highly relevant
Completeness: Complete Evidence
Overall Support: Evidence strongly supports the element
Possible Challenges: None

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

Evidence: In 2015, NCSC submitted the following evidence in support of peer review critical element 3.4, Evidence of Relations to Other Variables.

(a) Alignment between knowledge, skills, and abilities in assessment to student learning expectations for instruction. Chapter 2 Test Development; pp. 75-76 and Appendix 3-B, Study 2. Notes: This investigation affirmed that the targets for measurements provide information useful for tracking student progress in the CCSS and to teachers for providing instruction focused on academic expectations.

(b) Vertical coherence study. Chapter 2 Test Development; pp. 82-84 and Appendix 3-B, Study 5. Notes: This investigation addresses the extent to which assessment claims align with grade-level content and are useful for tracking progress. Results indicate measurement targets are appropriate and useful for these purposes.

In addition, the internal correlations between 2023 MSAA ELA and mathematics scores for grades 3–8 and HS are, in order, .86, .82, .85, .88, .80, .85, and .86. These correlations indicate a moderate to strong relationship between ELA and mathematics MSAA scores, reasonably consistent with correlations observed between grade-level state assessments and external measures (e.g., local interim assessments). They suggest that MSAA 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 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

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

States offer guidance to local districts for developing teacher professional development, as exemplified by the Arizona Department of Education'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**). While this evidence is moderately relevant, it's limited, and additional evidence, such as surveys of local education agencies (LEAs) to assess the implementation of MSAA-based professional development, would be valuable to further support the element.

Evidence: States provide guidance to local districts to promote and guide development of teacher professional development. For example, the Arizona Department of Education provides a document called How to Teach the State Standards to Students Who Take Alternate Assessments https://cms.azed.gov/home/GetDocumentFile?id=5866dbe1aadebe085c4de5b4

Furthermore, in the survey referenced in Assumption 2.1, There are 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: The evidence is highly relevant
Completeness: Moderate to substantial evidence
Overall Support: The evidence moderately supports the element. It's noteworthy that the demographics of the pilot survey may not entirely reflect the broader MSAA population.

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.

Assumption 3.1. Teachers use the MSAA 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: 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 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 scores useful for planning instruction.

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

Element 3.2.1. Teachers use MSAA 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).Additionally, teachers have access to MSAA teacher guides to gauge student achievement as well as for instructional planning.

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 scores and other MSAA-based information for planning instruction.

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 scores and other information useful for understanding what their child knows and can do.

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

Evidence: 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 http://www.azed.gov/assessments/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 correctly understand and interpret MSAA scores and other MSAA-based information to understand what their child knows and can do.

Element 4.1.2. Parents use MSAA 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 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 scores and other MSAA-based information to understand what their child knows and can do.

Assumption 4.2. Parents find MSAA scores and other information useful for understanding their child's progress from year to year.

Element 4.2.1. Parents understand and interpret correctly MSAA scores and other information to understand their child's progress from year to year.

Evidence: 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: 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 scores and other MSAA-based information to understand their child's progress from year to year.

Element 4.2.2. Parents use MSAA scores and other information appropriately to understand their child's progress from year to year and make decisions about their child's education and learning needs.

Evidence: 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 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 scores and other MSAA-based information to understand their child's progress from year to year.

Table P.1 Status of Relevance, Completeness, and Overall Support of the Evidence for All Four SIUs, Assumptions, and Elements

Element	Relevance of the Evidence		Completeness of the Evidence to the Element			Overall Support to the Element		
	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.								
1.1.1. MSAA content is aligned to the CCCs and grade-level standards.	X		X			X		
1.1.2. MSAA items are aligned to the CCCs.	X		X			X		
1.1.3. States have confirmed alignment of the MSAA to state content standards.	X		X			X		
1.1.4. MSAA items are aligned to the PLDs.	X			X			X	
1.2.1. Items require application of the KSAs of the targeted construct.	X			X		X		
1.2.2. Items are accessible to all students.	X			X		X		

continued

Element	Relevance of the Evidence		Completeness of the Evidence to the Element			Overall Support to the Element		
	Highly Relevant	Moderately Relevant	Complete Evidence	Moderate to Substantial Evidence	Limited Evidence	Strongly Support	Moderately support	Limited Support
1.2.3. Appropriate accommodations are provided to meet student needs.	X			X		X		
1.2.4. Scoring rubrics focus on construct-relevant aspects of student responses.	X			X		X		
1.2.5. Scaffolding is not a source of construct-irrelevant variance.	X			X		X		
1.2.6. Item rendering does not interfere with student access to test content.	X			X		X		
1.2.7. Platform does not interfere with student interaction with test content.	X			X		X		
1.2.8. Items are free of bias and sensitive issues.	X			X		X		
1.3.1. Test Administrators and School and District Coordinators understood and performed their roles properly.	X				X			X
1.3.2. Test security concerns were limited.	X			X			X	
1.4.1. MSAA scores and categorizations into performance levels are adequately reliable for their intended purpose.	X				X			X
1.4.2. Item characteristics support intended interpretations about all students who take the MSAA.	X		X			X		
1.4.3. Test characteristics, for Paths A, B, and C support intended interpretations about all students who take the MSAA.	X			X			X	
1.4.4. Scaling of the MSAA supports intended interpretations about all students who take the MSAA.	X			X			X	
1.4.5. Equating of MSAA test forms supports intended interpretations about MSAA students.	X			X			X	
1.4.6. State 1 covers a broad enough range of item difficulty and item cognitive complexity to route students into appropriate stage 2 tests.	X		X			X		
1.4.7. Routing into the stage 2 test level is appropriate for students.	X		X			X		
1.4.8. Stage 2 test levels are sufficiently separable and targeted toward different ranges of achievement for the MSAA students who are routed to those levels.	X		X			X		

continued

Element	Relevance of the Evidence		Completeness of the Evidence to the Element			Overall Support to the Element		
	Highly Relevant	Moderately Relevant	Complete Evidence	Moderate to Substantial Evidence	Limited Evidence	Strongly Support	Moderately support	Limited Support
1.5.1. Machine-scored items were scored accurately.	X		X			X		
1.5.2. Constructed-response item scoring training and monitoring procedures met industry standards.	X		X			X		
1.6.1. MSAA scores correlate as expected with other measures of student proficiency.	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.								
2.1.1. MSAA 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		X				X
2.1.2. MSAA 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		X			X		
2.1.3. The relationship between MSAA scores and external measures of student achievement and growth is as expected, compared to grade-level assessments and other alternate assessments.	X			X			X	
2.2 (Assumption) MSAA 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.								
3.1.1. Teachers find the performance level descriptors and their students' performance levels useful for planning instruction, especially students in performance levels 1 and 2.		X			X		X	
3.1.2. Teachers find their students' scale score information useful for planning instruction, especially students in levels 1 and 2.		X			X		X	
3.2.1. Teachers use MSAA scores and other information for planning instruction.		X			X		X	

continued

Element	Relevance of the Evidence		Completeness of the Evidence to the Element			Overall Support to the Element		
	Highly Relevant	Moderately Relevant	Complete Evidence	Moderate to Substantial Evidence	Limited Evidence	Strongly Support	Moderately support	Limited Support
Primary Intended Score Use 3 Parents understand and interpret correctly MSAA scores and other information to understand what their child knows and can do.								
4.1.1. Parents understand and interpret correctly MSAA scores and other information to understand what their child knows and can do.		X			X		X	
4.1.2. Parents use MSAA 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			X		X	
4.2.1. Parents understand and interpret correctly MSAA scores and other information to understand their child's progress from year to year.		X			X		X	
4.2.2. Parents use MSAA scores and other information appropriately to understand their child's progress from year to year and make decisions about their child's education and learning needs.	X				X		X	