Colorado Measures of Academic Success (CMAS)
Technical Report

2023–2024

Foreword

This technical report documents the evidence of reliability and validity to support test users in evaluating the intended purposes, uses, and interpretations of the test scores for the Spring 2024 administration of the Colorado Measures of Academic Success (CMAS) assessments. The evidence includes descriptions of the test design, development, and administration procedures; the student test results; and psychometric analyses including calibration, equating, and scaling to ensure that the test results can be compared across different test forms and administrations. The report adheres to industry best practices and follows the guidelines of the *Standards for Educational and Psychological Testing* (AERA et al., 2014).

The Colorado Department of Education's vision is to create an equitable educational environment where all students and staff in Colorado thrive. Their role is to improve student outcomes and ensure that students and families across Colorado have access to high-quality schools by providing actionable support to local educational agencies, implementing policy and legislation in an effective way, and sharing the experiences of local educational agencies and students.

Colorado Department of Education
201 East Colfax Ave., Denver, CO 80203
303-866-6600
https://www.cde.state.co.us/

Susana Córdova
Commissioner of Education

Rebecca McClellan
Chair, Colorado State Board of Education

Jared Polis
Governor

COLORADO
Department of Education

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The purpose of this technical report is to inform users and other interested parties about the development, administration, and technical characteristics of the Spring 2024 Colorado Measures of Academic Success (CMAS) assessments administered in English language arts (ELA), mathematics, and science to measure Colorado students' mastery of the Colorado Academic Standards (CAS) and comply with state and federal accountability requirements. The Colorado Spanish Language Arts (CSLA) assessments are also available to eligible students.

1.1. Assessment Overview

The CMAS assessments are Colorado's end-of-year standards-based assessments designed to measure students' achievement of the 2020 CAS each spring in grades 3–8 in ELA and mathematics and in grades 5, 8, and 11 in science. The CSLA forms of the ELA assessment are also administered to students with Spanish as their home language in grades 3 and 4 who meet established eligibility criteria. The CSLA forms serve as accommodated versions of the CMAS ELA assessments and are parallel and comparable to ELA in test design, scoring, and reporting.

The assessments are administered online as fixed forms, meaning all students receive the same set of operational items in a predetermined order but different embedded field test items depending on which test form they receive. The field test items do not count toward a student's score. Paper-based test forms are also available, along with a wide range of accessibility features for all students and accommodations for students with disabilities and multilingual learner students, including assistive technology forms, braille, large print, and text-to-speech (TTS). The tests are designed to be administered within a specified timeframe across two or three units and contain machine-scored selected-response (SR) and technology-enhanced (TE) item types and handscored constructed-response (CR) item types.

Student results are reported as an overall scale score and performance level, with a separate Reading scale score for ELA and science reporting category scale scores for Physical Science, Life Science, Earth and Space Science, and the Science and Engineering Practices (SEPs). The ELA, mathematics, and CSLA assessments have five performance levels, whereas science has four performance levels. Students in the top two performance levels met or exceeded the expectations of the CAS and are considered on track for the next grade level in the content area.

1.2. Background

The CMAS assessments were first administered in 2013–2014 for science and social studies and in 2014–2015 for ELA and mathematics, and the CSLA assessments were first administered in 2015–2016. Colorado developed the CMAS Mathematics and ELA assessments in collaboration with the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium, with Pearson taking over as the testing contractor in 2017–2018.[1] Pearson has been the testing contractor for the CMAS Science and Social Studies assessments and the CSLA assessments since their inception, although the social studies assessments have not been administered since 2014 for high school and 2019 for grades 4 and 7 due to legislative decisions.

[1] For information on the background of the consortium and the development and administration of the assessments, see prior years' technical reports at https://www.cde.state.co.us/assessment/cmas_coalt_techreport.

In 2017, the Colorado State Board of Education provided direction to the Colorado Department of Education (CDE) to decrease testing time. CDE explored the use of abbreviated versions of the prior years' test blueprints with the goal of decreasing testing time while retaining comparability to the previous CMAS ELA/CSLA and Mathematics assessments to maintain longitudinal trend data. Test forms based on the abbreviated blueprints were developed in Fall 2017 and administered beginning in Spring 2018.

The 2020 administration was cancelled because of the COVID-19 pandemic. In 2021, Colorado received a partial waiver of the federal assessment requirements from the U.S. Department of Education (USED) due to COVID-19 conditions in Colorado. The number of tests students were required to take was reduced, with alternating grades for mathematics and ELA and only students in grade 8 taking the science assessment.

In 2022, newly revised standards were implemented for mathematics, ELA, and science as a result of Senate Bill 212 (also known as CAP4K) implemented by Colorado in 2008 that required the State Board of Education to adopt content standards that prepare students for the 21st century workforce and for active citizenship upon receiving a high school diploma. It also required a revision to the CAS by July 1, 2018, and every six years thereafter. As such, the 2009/2010 CAS were reviewed and revised, resulting in the 2020 CAS. Minimal changes were made to the mathematics and ELA standards, and regular testing procedures resumed in Spring 2022. However, the science standards underwent a substantial update to keep up with the shift to the Next Generation Science Standards (NGSS; NGSS Lead States, 2013), resulting in the development of a new CMAS Science assessment.

Full implementation of the new three-dimensional science standards took place in 2021–2022. The new CMAS Science test was administered to all tested students for the first time in Spring 2022, which made it possible to test enough new content to allow for a robust item bank and to obtain a sufficient sample of students to conduct field test analyses. The Spring 2022 CMAS Science assessment reported percentile ranks only. Standard setting was conducted in Fall 2022, and overall scale scores and performance levels were reported beginning with the Spring 2023 administration.

1.3. Purpose of CMAS

The CMAS assessments serve multiple purposes, including informing parents/guardians and educators about individual student achievement of the grade-level CAS and allowing comparisons to other students across the state. Results are intended to provide one measure of a student's academic progress relative to the CAS and should be considered alongside other achievement information available locally. State assessment data also help inform the state's school and district accountability system, including assigning performance ratings to schools and districts. State assessment results may also be a component of educator evaluation.

In general, CMAS is a source of data that (a) may be used as a prompt for further investigation at the student, classroom, school, and district levels; (b) supports districts/schools in reviewing and developing goals for the performance of their students, including subgroups; (c) may indicate that a review of programs, curricula, materials, and/or scope and sequence may be appropriate; and (d) may inform the evaluation of district/school approaches.

Assessment results also support a range of data-driven stakeholder conversations, activities, and decisions such as school selection, program evaluation, investigative research, and policy/legislation formation and review. For example, educators can use the test scores to plan for further instruction and curriculum development and to report progress to parents/guardians. The results can also be used as one factor in making administrative decisions about program effectiveness, teacher effectiveness, class grouping, and needs assessment or for research purposes and informing community and organization efforts.

1.4. Testing Requirements

All public schools in Colorado are required by state law to administer the standards-based summative assessments each year in the specified content areas and grade levels to comply with the federal accountability requirements as stated in the Every Student Succeeds Act of 2015 (ESSA). As a requirement of Colorado School Law C.R.S. §22-7-1006.3 (4) (II)(b), students with Spanish as their home language in grades 3 and 4 who meet established eligibility criteria may take the CSLA forms of the ELA assessment.

The CMAS assessments are intended to be taken by all students enrolled in public schools, except for some students with the most significant cognitive disabilities who take the Colorado Alternate (CoAlt) assessment as determined by the student's Individualized Education Plan (IEP) team. Exempt students include those that are excused as a result of Colorado legislation C.R.S. §22-7-1013 (8) (a-c) passed in 2015 that allows for parents/guardians to excuse their child(ren) from testing. However, every student, regardless of ability or language background, must be provided the opportunity to demonstrate their content knowledge through the state assessments.

Colorado legislation (C.R.S. §22-7-1006.3 (1) (d)) also requires that a paper-based version be available for all online assessments that may be used by local educational providers for their students. The comparable paper-based forms may also be administered to students with disabilities and multilingual learner (ML) students as appropriate. Multilingual learners for assessment purposes are students with a home language other than English who are designated as not English proficient or limited English proficient (NEP/LEP) by an English language proficiency assessment or screener.

ML students in their first year in the United States are exempt from the ELA assessment, but ML students in grades 3 and 4 designated as NEP whose native language is Spanish and who have received language arts instruction in Spanish during the current school year are required to take the CSLA assessment. Students with disabilities and ML students may take the CMAS assessments with or without accommodations that do not change the construct of the assessment. Accommodations are determined based on classroom experience and educational team decisions.

1.5. Assessment Development Partners

CMAS assessment activities were conducted collaboratively by CDE, the Colorado educator community, and Pearson, with input and advice from the Colorado Technical Advisory Committee (TAC), as shown in Table 1.1. Each contributor plays a vital role in ensuring that the assessments yield valid and reliable test results. Educator participation in the test development process is critical to ensuring that the assessments are aligned to the CAS, are appropriate for Colorado students at the assessed grade level, and are free from bias and sensitivity issues, and recommendations from the TAC have been reviewed, addressed, and incorporated into the assessments.

Table 1.1. Assessment Development Partners

Organization/Group	Roles and Responsibilities
Colorado Department of Education (CDE)	• The administrative arm of the State Board of Education responsible for implementing state and federal education laws • Works closely with Colorado school districts, educators, community stakeholders, and test development partners to develop and administer the state assessments, focusing on creating assessments that serve students, schools, districts, and the community while complying with state and federal legal requirements • Works closely with Pearson on each facet of the assessment, with CDE serving as the ultimate approver of services and products provided
Colorado Educator Community	• Create assessment items aligned to the CAS, with items that successfully move through the entire item development process eventually appearing on the operational assessments • Review items to ensure content alignment and identify potential bias and sensitivity concerns before items are field tested • Participate in rangefinding to review student responses to field tested CR items and define the score point ranges for the scoring rubrics used to score student responses • Participate in data review to review field tested items with statistical parameters outside of normal ranges to determine if the items are acceptable for inclusion in the operational item bank
Pearson	• Primary contractor responsible for the end-to-end assessment cycle services and products • Works closely with CDE throughout the CMAS and CoAlt Science assessment development and administration processes, including item and test development, forms creation, enrollment, packaging and distribution, test delivery, scoring, customer service, standard setting, scoring, score reporting, and psychometric services
Tri-Lin Integrated Services, Inc.	• Subcontractor to Pearson responsible for CSLA content and test development, including passage development, item development, and test form construction
Technical Advisory Committee (TAC)	• A group of psychometric, assessment, and special populations experts who provide high-level consulting and expert advice regarding validity and reliability issues on topics such as blueprint design, scaling and equating, mode comparability, scoring, reporting, alignment study feedback, peer review, and standard setting • Included the following members during the 2024 assessment cycle: – Dr. Elliot Asp, Senior Partner, The Colorado Education Initiative – Dr. Jonathan Dings, Executive Director of Student Assessment and Program Evaluation, Boulder Valley School District – Dr. Michael Kolen, Psychometric Consultant – Dr. Suzanne Lane, Professor, University of Pittsburgh – Dr. Martha Thurlow, Director, National Center on Educational Outcomes – Dr. Jon Twing, Chief Scientist, HumRRO

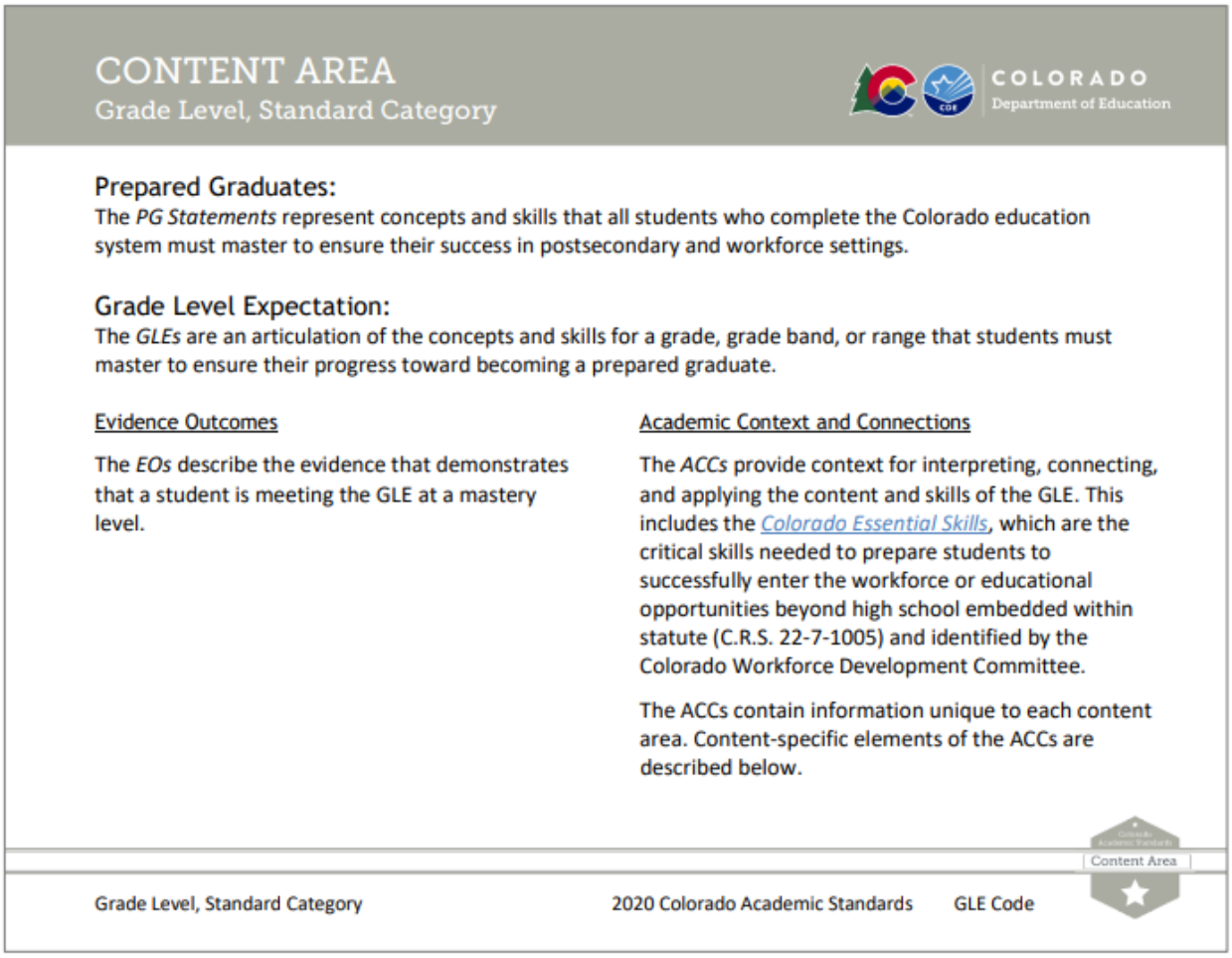
Chapter 2: Test Design

2.1. Colorado Academic Standards

The CMAS assessments are standards-based tests designed to measure what students should know and be able to demonstrate at the end of each grade or grade band based on the 2020 CAS located at the following links for each content area. The CAS for all content areas include the components in Figure 2.1.

- 2020 Mathematics Standards: http://www.cde.state.co.us/comath/statestandards
- 2020 Reading, Writing, and Communicating Standards: http://www.cde.state.co.us/coreadingwriting/statestandards
- 2020 Science Standards: https://www.cde.state.co.us/coscience/statestandards

Figure 2.1. How to Read the Colorado Academic Standards



The 2020 CAS for Mathematics and ELA had minimal changes compared to the previous 2009/2010 standards, whereas the 2020 CAS for Science underwent significant changes to be based on the NGSS.[2] The NGSS were guided by *A Framework for K–12 Science Education* (National Research Council, 2012) and designed to reflect more recent research and thinking in science education. The 2020 CAS for Science represent what all Colorado students should know and be able to do in science based on their pre-k–grade 12 science education.

The new science content standards are considered three-dimensional in that they incorporate Disciplinary Core Ideas (DCIs), Science and Engineering Practices (SEPs), and Crosscutting Concepts (CCCs). The DCIs encompass the content that occurs at each grade and provides the background knowledge for students to develop sense-making around phenomena in the three standards of Physical Science, Life Science, and Earth and Space Science. The DCIs are as follows[3]:

- Physical Science: Students know and understand common properties, forms, and changes in matter and energy.
    - PS1: Matter and its interactions
    - PS2: Motion and stability: Forces and interactions
    - PS3: Energy
    - PS4: Waves and their applications in technologies for information transfer

- Life Science: Students know and understand the characteristics and structure of living things, the processes of life, and how living things interact with each other and their environment.
    - LS1: From molecules to organisms: Structures and processes
    - LS2: Ecosystems: Interactions, energy, and dynamics
    - LS3: Heredity: Inheritance and variation of traits
    - LS4: Biological evolution: Unity and diversity

- Earth and Space Science: Students know and understand the processes and interactions of Earth's systems and the structure and dynamics of Earth and other objects in space.
    - ESS1: Earth’s place in the universe
    - ESS2: Earth’s systems
    - ESS3: Earth and human activity

The SEPs describe how scientists investigate and build models and theories of the natural world or how engineers design and build systems. They reflect science and engineering as they are practiced and experienced. There are eight SEPs:

[2]A summary of all the changes made to the standards are available on the CDE website for mathematics at https://www.cde.state.co.us/comath/2020cas-ma-changes, for ELA at https://www.cde.state.co.us/coreadingwriting/2020cas-rw-changes, and for science at https://www.cde.state.co.us/coscience/2020cas-sc-changes.

[3]Adaptation of the NGSS occurred by not adopting the fourth standard of Engineering, Technology, and Applications of Science (although engineering is still incorporated within the SEPs).

1. Asking questions (for science) and defining problems (for engineering)
2. Developing and using models
3. Planning and carrying out investigations
4. Analyzing and interpreting data
5. Using mathematics and computational thinking
6. Constructing explanations (for science) and designing solutions (for engineering)
7. Engaging in argument from evidence
8. Obtaining, evaluating, and communicating information

CCCs cross boundaries between science disciplines and provide an organizational framework to connect knowledge from various disciplines into a coherent and scientifically based view of the world. They build bridges between science and other disciplines and connect the DCIs and SEPs throughout the fields of science and engineering. There are seven CCCs:

1. Patterns
2. Cause and Effect
3. Scale, Proportion, and Quantity
4. Systems and System Models
5. Energy and Matter
6. Structure and Function
7. Stability and Change

The CMAS Science assessment is given in grades 5, 8, and 11. Consistent with the standards, the grade 5 assessment assesses the grade-level standards. Because the science standards are articulated by grade band at the middle school and high school levels rather than grade levels, the grade 8 assessment assesses all middle school science standards, and the grade 11 assessment assesses all high school science standards.

2.2. Test Frameworks and Blueprints

Concepts and skills identified in the CAS are the basis for the CMAS assessments. The CMAS frameworks list the number of score points for each subclaim and standard area that appear on the assessments and percent of the test represented by that subclaim or standard. The frameworks specify the Evidence Outcomes (EOs) from the CAS that are included on the assessments. The mathematics and ELA frameworks continue to use Evidence Statements (ES) developed in collaboration with PARCC that describe the knowledge and skills an assessment item/task elicits from students. Together, the CMAS frameworks and ES provide the foundation for ensuring that the full range and depth of the standards are assessed. CDE incorporated feedback from content experts and educators throughout the state to create the final versions of the frameworks. The frameworks and ES are both available on the CDE website at https://www.cde.state.co.us/assessment/cmas_testdesign.

The test blueprints take the frameworks a step further by specifying the number of test items by Prepared Graduate (PG) Statement, Grade-Level Expectation (GLE), EO, item type, and cognitive complexity. The specificity of the test blueprints ensures that the assessments cover the breadth of the content indicated by the CAS within the associated grade or grade band. Appendix A presents the high-level test blueprints that summarize the percentage of score points on each test for each claim and subclaim on each assessment as shown in the frameworks. The most recent versions of the ELA and mathematics blueprints were developed in 2017–2018, while new test blueprints were created for science in 2021–2022.

2.2.1. ELA and Mathematics

In 2017, the State Board of Education provided direction to CDE to decrease testing time. CDE began exploring the use of abbreviated versions of the prior years' test blueprints with the goal of decreasing testing time while retaining comparability to the CMAS Mathematics and ELA assessments previously administered in Colorado to maintain longitudinal trend data. Therefore, with the intent to reduce testing time, the 2018 blueprints were a proportionate abbreviation of the 2017 forms. CDE and Pearson collaborated in designing the CMAS subject- and grade-specific blueprints for ELA and mathematics in 2017–2018. The blueprints were designed to measure the same constructs as, and provide content comparability to, the previous year's assessments. Eligible content continued to reflect the CAS and ES used in prior years.[4]

2.2.2. Science

Pearson worked with Achieve, a nonprofit education organization that leads the effort to help states make college and career readiness a priority for all students, during the initial development of the new science assessment. Achieve provided background on how other states were approaching the new three-dimensional science standards and assessments and advice on how to proceed with cognitive complexity, blueprints, and reporting. With guidance from Achieve, Pearson, CDE, and Colorado educators collaborated in designing the science blueprints in a workshop held from November 6–7, 2019, in Denver. An effort was made to involve educators who were from areas representative of the entire state of Colorado (in terms of geographic location, gender, and race) and familiar with the 2020 CAS, related three-dimensional science instruction, and the assessment interaction and demonstration of achievement of the CAS of different groups of students, including students with disabilities and ML students. The blueprints were reviewed on October 14, 2021, by the TAC.

Results from the Spring 2022 test administration showed that items at the end of the test units were often left unanswered on the grade 11 CMAS Science assessment, indicating that high school students had difficulty finishing the test in the time given. Therefore, after review by CDE and Pearson psychometrics, a proportional reduction was made to the grade 11 blueprint to prevent speededness. The proposed blueprint reduction and test timing data was reviewed by the TAC on August 30, 2022. Grades 5 and 8 did not show a similar concern with speededness, so the length of these tests was not changed.

[4] For more information about the transition and abbreviated assessments, see the 2017–2018 CMAS Mathematics and ELA technical report on the CDE website at https://www.cde.state.co.us/assessment/cmas_coalt_techreport.

2.3. Claims and Subclaims

Student performance on the CMAS assessments is reported at the overall content area level as a scale score and performance level. Their performance is broken down even further at the claim and subclaim levels. The mathematics subclaims provide information on a student's achievement on grade-level math skills and concepts, as well as reasoning and modeling based on both grade-level and securely held knowledge of the skills and concepts from the previous grade level. The Reading and Writing claims for ELA provide information on a student's achievement in reading and comprehending a range of sufficiently complex texts independently. The subclaims are intended to provide more granular information about student demonstration of the knowledge and skills within the content area as reflected in the CAS.

Table 2.1 presents the content reflected in each subclaim by content area. The mathematics score is a composite of the four subclaims (Major Content, Supporting Content, Mathematical Reasoning, and Modeling and Application). The Reading score is a composite of the three reading subclaims (Reading: Literary Text, Reading: Informational Text, and Reading: Vocabulary, and the Written Expression subclaim that measures reading), and the Writing claim is a composite of the two writing subclaims (Writing: Written Expression and Writing: Knowledge and Use of Language Conventions). The science score is a composite of the three standards (Physical, Life, and Earth and Space Science), as well as an SEP score.

Table 2.1. Subclaims

Content Area	Subclaim	Description
Mathematics	Subclaim A: Major Content	Students solve problems involving the Major Content of the grade level with connections to the Standards for Mathematical Practice.
	Subclaim B: Additional & Supporting Content	Students solve problems involving the Additional and Supporting Content of the grade level with connections to the Standards for Mathematical Practice.
	Subclaim C: Expressing Mathematical Reasoning	In connection with content, the student expresses grade/course-level appropriate mathematical reasoning by constructing viable arguments, critiquing the reasoning of others and/or attending to precision when making mathematical statements.
	Subclaim D: Modeling & Application	In connection with content, the student solves real-world problems with a degree of difficulty appropriate to the grade/course by applying knowledge and skills articulated in the standards for the current grade/course (or for more complex problems, knowledge and skills articulated in the standards for previous grades/courses), engaging particularly in the Modeling practice, and where helpful making sense of problems and persevering to solve them, reasoning abstractly and quantitatively, using appropriate tools strategically, looking for the making use of structure, and/or looking for and expressing regularity in repeated reasoning.
ELA	Reading: Literary Text	Students read and analyze fiction, drama, and poetry.
	Reading: Informational Text	Students read and analyze nonfiction, history, science, and the arts.
	Reading: Vocabulary	Students use context to determine what words and phrases mean.
	Writing: Written Expression	Students compose well-developed writing using details from what they have read.
	Writing: Knowledge and Use of Language Conventions	Students demonstrate knowledge of conventions and other important elements of language.

Content Area	Subclaim	Description
Science	Physical Science	Students know and understand common properties, forms, and changes in matter and energy.
	Life Science	Students know and understand the characteristics and structure of living things, the processes of life, and how living things interact with each other and their environment.
	Earth and Space Science	Students know and understand the processes and interactions of Earth's systems and the structure and dynamics of Earth and other objects in space.
	Science and Engineering Practices (SEPs)	The SEPs describe how scientists investigate and build models and theories of the natural world or how engineers design and build systems. They reflect science and engineering as they are practiced and experienced.

2.4. Cognitive Complexity

All mathematics and ELA items are tagged with a cognitive complexity level of high, moderate/medium, or low, as described in Table 2.2.

Table 2.2. Mathematics and ELA Cognitive Complexity Levels

Content Area	High	Moderate/Medium	Low
Mathematics	• Significant shift from previous content • Open ended, sophisticated reasoning, critiquing, modeling • Single/multi-part that requires more evidence from the student	• Moderate shift into new content • Moderately scaffolded, some choice in approach • Single/multi-part, multi-step, moderate reading load	• Low shift from previous content • Very scaffolded, rote, recall, recognize • Single part, one step with low reading load
ELA	Items require synthesis of ideas and details **across multiple texts or ideas** (can be single passage). For example, items may require students to construct the main idea or theme that is common across multiple texts, especially multiple texts that are not closely related in theme and/or genre.	Items require **analysis of ideas and details** across multiple sections **in a single text**. It requires more close analytic reading than low complexity items. For example, identifying the main idea or theme of a text may require inferring the main or theme or integrating ideas and details from several locations in the text.	Items require students to identify **a single idea or detail in a text** (e.g., identifying a term or phrase using context). It requires students to recall, observe, question, or represent facts or simple skills or abilities.

Science transitioned away from Depth of Knowledge (DOK) in 2021–2022 with the adoption of the new science standards. From Achieve:

> *As states and districts develop new assessment systems, they need support for developing assessments that balance the vision and integrity of multi-dimensional standards with ensuring that they are sensitive to varying levels of student performance. This... (requires a) ...new approach to capturing and communicating the complexity of summative assessment items and tasks designed for three dimensional standards that can be used to ensure that all learners can make their thinking and abilities visible without compromising the rigor and expectations of the standards* (Achieve, 2019, p. 1).

The CMAS Science assessment now uses a cognitive complexity framework that examines items via three criteria, as summarized in Table 2.3 and presented in Appendix B. Phenomenon in the stimulus material is examined separately for its own cognitive complexity.

Table 2.3. Science Cognitive Complexity Criteria

Criterion	Description
Item Dimensionality	Item alignment to one, two, or three dimensions: • Content of EO (Disciplinary Core Idea DCI) • Science and Engineering Practice (SEP) of EO • Cross Cutting Concept (CCC) of EO Items aligned to a single dimension only are **not** acceptable for CMAS Science.
Scaffolding/Support	The more guidance and structure the item provides the student, the lower the cognitive load required. The matrix categorizes scaffolding/support into three levels: heavy, moderate, and minimal. Heavy refers to a specific, step-by-step process is given, and the student merely needs to follow that process to supply the answer. Moderate and minimal provide increasing degrees of freedom to make choices on the part of the student and require an increasing degree of initiative to make those choices.
Sensemaking	Fundamental to the approach of three-dimensional standards is student use of the dimensions to make sense of scientific phenomena. Some degree of sensemaking is required for all CMAS Science items. A sensemaking situation is one in which students (1) are provided material without obvious ties/connections to content (e.g., language of the standard) and (2) use their knowledge of the standard to explain what they see in the material.

2.5. Item Types

CMAS Mathematics and Science contain selected-response (SR), technology-enhanced (TE), and constructed-response (CR) items. Mathematics also contains fill-in-the-blank (FIB) items. The CMAS ELA/CSLA assessments are passage-based with a combination of literary and informational passages and contain SR, TE, and prose constructed-response (PCR) items. Multiple passages may be used to respond to some items.

For the ELA PCRs, students receive a prompt, respond to reading items, and write an extended response. It is then scored on a multi-trait rubric, as provided on the CDE website at https://www.cde.state.co.us/assessment/cmas_testdesign. The ELA PCRs include three task types: literary analysis, research simulation, and narrative writing. Because it is administered on paper, CSLA forms contain SR, paper-based TE, and PCR items. The CSLA paper-based TE items are developed to have similar item formats and scoring rules to the paper-based versions of TE items developed for CMAS ELA.

All mathematics items are aligned to both an ES and an EO. The ES are grouped into three types to ensure that the full range and depth of the standards are assessed:

- Type I items:
  - Assess a specific EO, a specific part of an EO, or multiple EOs
  - Subclaims A and B
  - 1- or 2-point items (grades 3–8) and 4-point items (grades 6–8)
  - SR, TE, and FIB items
  - Calculator (grades 6–8) and non-calculator (grades 3–8)

- Type II items (reasoning):
    - Assess a specific type of mathematical reasoning and a specific scope in the EOs to reason about
    - Subclaim C
    - 3- or 4-point items
    - SR, TE, FIB, and CR parts; all items have at least one CR part
    - Calculator (grades 6–8) and non-calculator (grades 3–5)

- Type III items (modeling):
    - Assess a specific type of mathematical modeling and a specific scope in the EOs to model about
    - Subclaim D
    - 3- or 6-point items
    - SR, TE, FIB, and CR parts; all items have at least one CR part
    - Calculator (grades 6–8), non-calculator (grades 3–5)

The CMAS Science assessment is divided into item sets that present phenomenon-based scenarios as either interactive science simulations or static stimuli, followed by associated standalone items or clusters of items related to the simulation or scenario. A phenomenon is an observable event that students can use the three dimensions (DCI, SEP, and CCC) to explain or make sense of. Separate standalone items are also included that are not associated with a stimulus to target a small number of CAS not represented in the scenarios. The items are either 1-point SR, 1-point TE, or 2-point CR item types.

2.6. Test Units

Each assessment was composed of two or three units with embedded field test items to allow the assessments to be administered in a reasonable timeframe, as shown in Table 2.4.

Table 2.4. Test Units

Grades	Mathematics	ELA	Science
3–5	Units 1–3: 65 minutes Total time: 195 minutes	Units 1–3: 90 minutes Total time: 270 minutes	Units 1–3: 80 minutes Total time: 240 minutes
6–8	Units 1–3: 65 minutes Total time: 195 minutes	Units 1–3: 110 minutes Total time: 330 minutes	Units 1–3: 80 minutes Total time: 240 minutes
11	N/A	N/A	Units 1–2: 50 minutes Total time: 100 minutes

Chapter 3: Item Development

The CMAS item development process results in a diverse bank of items that align to the CAS. All items are developed with the intention of being administered on multiple testing platforms, including online, online-accommodated, and paper-based assessments. The item writing process is a tiered, inter-related process that began with the development of the test blueprint for each grade level within each content area, followed by creating the item development plan used to forecast the targeted number of items and associated stimuli across ESs or EOs needed to create a robust item bank.

Once written, all newly developed items go through multiple rounds of review, including contractor, CDE, and Colorado educator content, bias, and data reviews. Table 3.1 presents the item development activities for the items field tested on the Spring 2024 assessments. The ELA passage review included five windows where educators reviewed batches of passages independently.

Table 3.1. Item Development Activities

Event	Date(s)
ELA Passage Reviews	Text Review 1: January 17–23, 2023
	Text Review 2: January 31 – February 6, 2023
	Text Review 3: February 14–20, 2023
	Text Review 4: February 27 – March 3, 2023
	Text Review 5: March 15–21, 2023
Sim Storyboard Review (Science)	January 24, 2023
IWW Training (Science)	January 24–26 and March 8–10, 2023
IWW Training (ELA)	February 7–9, 2023
IWW Trainings (Math)	March 8–10, 2023
Content and Bias Review (Science)	July 18–21, 2023
Content and Bias Review (Math)	July 19–20, 2023
Content and Bias Review (ELA)	July 25–27, 2023
Data Review (CSLA)	July 31, 2024
Data Review (Science)	August 1–2, 2024
Data Review (ELA)	August 14–15, 2024
Data Review (Math)	September 11–12, 2024

As part of the test construction process, a selection of the proposed set of operational items are refreshed, as illustrated in Table 3.2. Therefore, a portion of the operational items have been used operationally on a previous CMAS form, while the remaining items are refreshed using Colorado-developed field test items. All items were reviewed by Colorado educators. (Please note that the Spring 2024 CMAS Grade 11 Science assessment included a set of core items held constant from 2023 to 2024 with a proportional reduction in length of the form.)

Table 3.2. Refresh Rates

Content Area	Item Type	Refresh Rate Minimum Targets	Refresh Rate Maximums
Mathematics	Type I: 1-point	25%	50%
	Type I: 2- and 4-point	40%	60%
	Type II	33%	67%
	Type III	50%	50%
ELA	Selected Response	50%	–
	Short Constructed Response	50%	–
	Extended Constructed Response	50%	–
	Passage Sets	50%	–
Science	Regular Cluster	33%	–
	Mini Cluster	50%	–
	Overall Items	33%	66%

3.1. Content Management Tool

Pearson's proprietary software, ABBI (Assessment Banking and Building solutions for Interoperable assessments), is used to support the test development process from initial content authoring through the review cycles. ABBI is the authoritative source for all content, data, and functionality for all CMAS system components. It serves as the repository where the item bank is housed, item revisions are catalogued, and items and item metadata are uploaded and revised by assessment specialists. Items can be moved into various statuses, each representing a step in the item development process. The items and associated stimuli are tracked, and revisions are recorded from creation through retirement in a secure environment.

Custom development reports can be generated out of ABBI, which allows users to generate Excel reports that capture metadata (e.g., unique item number, ES, task type, cognitive complexity, associated stimulus, item status, item statistics, and comments) useful for analyzing the item bank. ABBI is the source of reference for how and when changes to the item and the metadata have been implemented.

3.2. Item Development Plan

An item development plan for each content area and grade is created at the beginning of each item development cycle to determine the number of items, passages, and science cluster stimuli needed to construct the assessments based on the blueprint requirements, with development targets that address any task model, passage type, ES, EO, item/task type, and cognitive complexity shortages. To accomplish this, the item bank is analyzed, and the ES, EO, task type, and cognitive complexity gaps are identified so a variety of item types aligning to the ES, EOs, and the corresponding CAS can be created.

3.3. ELA Passage Development

Item development for CMAS ELA begins with the selection of literary and informational texts, whereas CSLA passages are commissioned by Tri-Lin, either in-house or by professional passage writers, due to the availability of appropriate passages and challenges with acquiring permissions. The number and types of needed passages are determined by the test construction specifications, a gap analysis of the pool of available passages, and the item development plan. Contractor assessment specialists train passage searchers to find (or write for the CSLA items) relevant and rich texts that permit a range of content to be developed.

Passage searchers and writers submit the passages for the contractor assessment specialists to review and evaluate using approved criteria, including adherence to the cognitive demand, relevance, and purpose of the test and the appropriate use of graphics to improve text comprehension. Test passages are analyzed and rated for text complexity. The assessment specialists check the passages for clarity, correctness of language, appropriateness of language for the grade level, and adherence to the style guidelines.

Accepted passages are presented to CDE for review. Once approved, the passages are reviewed for content and bias by committees of educators from throughout the state representing a variety of student populations, including students with disabilities and ML students. Passages accepted by both CDE and the educator committees are then used for item writing.

3.4. Science Scenario Development

Item development for science begins with the composition of the interactive simulations (SIMs) and cluster stimuli. The number and types of needed simulations and cluster stimuli are determined by the test construction specifications, a gap analysis of the pool of available SIMs and stimuli, and the item development plan. The topics are researched for suitability of science content, alignment to the standards, and grade-level appropriateness. The SIMs and cluster stimuli follow slightly different paths through the development process, but both include multiple steps of review by assessment specialists for adherence to cognitive complexity requirements, relevance to standards, purpose of the test, and the appropriate use of graphics and or animations. Pearson checks all stimulus text for scientific accuracy, clarity, correctness of language, appropriateness of language and science concepts for the grade level, and adherence to the style guidelines.

Simulation ideas are presented to CDE in the form of storyboards illustrating the intended virtual interaction, along with suggested EOs that the simulations address. Once CDE provides feedback, revised storyboards are reviewed by committees of educators from throughout the state representing a variety of student populations, including students with disabilities and EL/ML students. The SIMs are then developed into animated interactions and reviewed by CDE, after which items are written to a variety of EOs, either internally or by educators.

Cluster stimuli are proposed as topics to CDE and then developed into drafts based on CDE feedback. Drafts are refined by Pearson with CDE input and presented to educators for review and item writing using the same criteria used for the SIMs.

3.5. Item Writing

Item writer workshops (IWWs) were conducted with Colorado educators from across the state representing a variety of student populations, including students with disabilities and students with limited English proficiency. CSLA item writers are proficient in written academic Spanish and begin developing CSLA items after receiving training. The educators are given item writing assignments and develop a variety of items across task types, ES, and EOs. The item writers work with Pearson and/or Tri-Lin assessment specialists when clarification is needed for CSLA items. CDE content specialists are also present to assist as needed. Item writers use the ESs and EOs; the CAS; secure item specification documents, including item-writing guidelines (universal design guidelines, bias and sensitivity guidelines, and editorial guidelines); and an item writing checklist to guide them in completing their assignments.

All item writers author the items in ABBI, where Pearson or Tri-Lin assessment specialists complete their initial review. The assessment specialists review and suggest revisions to the items and metadata for the item authors, who then make the revisions and resubmit the items within ABBI.

3.6. Item Review

3.6.1. Internal Review

Pearson and Tri-Lin assessment specialists evaluate each newly developed item for content correctness; grade appropriateness; and ES, EO, CAS, and cognitive complexity alignment, focusing on the quality of the items, adherence to the principles of universal design, cognitive demand, relevance to the purpose of the test, and appropriateness of graphics. Research librarians perform additional fact checking to ensure accuracy. Pearson and Tri-Lin copy editors check items for clarity, correctness of language, appropriateness of language for the grade level, adherence to style guidelines, and conformity with acceptable item-writing practices.

All human-scored CR items are reviewed for their scorability by a Scoring Services director, and items and/or scoring rubrics with score points deemed "difficult to score" are revised in collaboration with the assessment specialist(s). All fill-in-the blank equation editor (FIBEE) items and machine-scored CR mathematics items and their rubrics are reviewed for their scorability by the Math Reasoning Engine (MRE) team, and items and/or scoring rubrics are revised in collaboration with the assessment specialist(s) before the MRE automated scoring is applied to the items.

Pearson and Tri-Lin assessment specialists also perform a universal design review to assess item accessibility irrespective of diversity of background, cultural tradition, and viewpoints; to evaluate changing roles and attitudes toward various groups; to review the role of language in setting and changing attitudes toward various groups; to appraise contributions of diverse groups (including ethnic and minority groups, individuals with disabilities, and women) to the history and culture of the United States and the achievements of individuals within these groups; and to edit for inappropriate language usage or stereotyping with regard to sex, race, culture, ethnicity, class, disability, or geographic region. The universal design review also includes reviewing items for potential bias to ensure that all items are fair and all students would have an equal opportunity to demonstrate achievement regardless of their gender, ethnic background, religion, socioeconomic status, disability, or geographic region. Items are also reviewed for visual bias, accessibility for students with disabilities, and convertibility to braille and text-to-speech.

Once the internal reviews are complete, each item's status is updated in ABBI and a lead assessment specialist conducts a final content review. Item statuses are updated in ABBI upon approval, and items are presented to CDE for review. Adhering to these processes ensures that each Colorado item measures the ES or EO and standard, is content- and grade-appropriate, is factually accurate, has appropriate answers and distractors, is accessible to all populations required to take the assessments, is free from any bias, and follows the Colorado style guidelines.

3.6.2. CDE Review

CDE reviews items in ABBI to ensure that the content is correct, the alignment is sound, the cognitive complexity is appropriate, the language and content are grade-appropriate, the graphics are clear and relevant to the item, and the content is free of bias/sensitivity issues. Once complete, CDE alerts Pearson or Tri-Lin. CDE's comments and determinations regarding the status of the items are recorded in ABBI, as indicated below:

- Items marked "Accept" need no more revisions and are ready for external Colorado educator content and bias reviews.
- Items marked "Accept with Edits" are revised per CDE's feedback and re-reviewed by the internal review team if necessary. These items are then reviewed by CDE again, reconciled with the assessment specialists, and deemed either "Accept" or "Reject."
- Items marked "Reject" are rejected and given a status of "Do Not Use" in ABBI. These items are either rewritten or replaced with items written by an assessment specialist. In either case, the items go through the same rigorous review process as newly developed items.

3.6.3. External Content and Bias Review

All items that pass the internal and CDE reviews are brought to external content and bias committees comprised of Colorado educators from across the state with diverse backgrounds and experience working with diverse learners (e.g., based on gender, race/ethnicity, income, and geography), standards and content expertise, and special population expertise (i.e., students with disabilities and EL/ML students). For science, educators are also selected based on their experience in the domain they are reviewing. For the accommodated CSLA items, an effort is made to involve educators who teach ML students, are familiar with the instruction and needs of the students in an English language development program that uses native language instruction, and are proficient in written Spanish.

The purposes of these educator reviews are to (a) ensure that the items are properly aligned to the CAS, accurately measure the intended content, and are grade-appropriate; and (b) identify any potential bias or stereotypes in the items. Separate committees are convened for each content area, as well as for the accommodated CSLA items. The meetings are conducted either in person or virtually and include group training on the expectations and processes of each meeting, followed by breakout groups by content area and grade where additional training is provided. The committee members are trained and instructed to verify that each item and stimulus

- displays and functions correctly in the online testing platform;
- aligns to the ES and/or EO;
- uses clear, unambiguous, and grade-appropriate language;
- avoids construct-irrelevant complex sentence structure and uses everyday words to convey meaning when vocabulary is not part of the tested construct;
- has one correct answer (depending on the item type) and contains plausible distractors that represent feasible misunderstandings of the content (depending on the item type);
- represents the range of cognitive complexities and includes challenging items for students performing at all levels;
- is appropriate for students in the assigned grade in terms of reading level, vocabulary, interest, and experience;

- has scoring guidelines that capture exemplar responses at each score point for CR items;
- includes appropriate and clear graphics/art/photos that are relevant to the item and accessible to all testing populations;
- is free of ethnic, gender, political, and religious bias;
- avoids construct-irrelevant content that may unfairly advantage or disadvantage any student subgroup; and
- considers access issues at the time of item writing (e.g., determine how students with visual disabilities would access items with needed visuals/graphics/animation).

The committees make one of three recommendations on every item: "Accept," "Accept with Edits," or "Reject." Following the educator meetings, CDE, Pearson, and Tri-Lin review committee comments, reconcile proposed edits, and finalize item outcomes. ABBI is updated to reflect the edits and outcomes. The approved items, passages, and simulations/clusters are then made ready for inclusion on the spring operational forms as embedded field test items.

3.7. Data Review

After item development is complete, selected items are placed on the operational assessments in embedded field test positions. The goal of field testing is to allow for the evaluation of the quality of the items through a review of item performance data to determine their inclusion in the operational item pool. To accomplish this, psychometricians perform statistical analyses on the field tested items following their administration in a field test environment to evaluate their quality.

Table 3.3 presents the statistical flags applied to the field tested items. Classical statistics include item means (*p*-values), item-total correlations/point biserials, and distribution of responses across answer options or score points, depending on the item type. Differential item functioning (DIF) analyses are conducted on various subgroups (gender, ethnicity, free and reduced lunch, IEP, and MLs) using Mantel–Haenszel Delta DIF statistics (Dorans & Holland, 1992). The same analysis methods are used for CSLA items, but the DIF analyses are conducted by gender only due to the population of students taking the form. Classification rules derived from National Assessment of Educational Progress (NAEP) guidelines (Allen et al., 1999) were used to classify items as having either negligible, moderate, or significant DIF. Items are then flagged based on the criteria in Table 3.3, and flagged items are taken to a data review meeting where a committee of educators reviews the flagged items and their statistics along with student performance data.

Table 3.3. Item Statistical Flagging Criteria

Statistic	Criterion	Possible Indication
P-value	< 0.1 or > 0.9	Very difficult or easy item
Item-total correlation	< 0.15	Poorly discriminating item
Distractor item-total correlation (SR only)	> 0.0	Possible miskey*
Score point percentage (multi-point items only)**	<1%, >50%, or >60%	Very few students or many students got a certain score
Differential item functioning (DIF)***	B, C	Item could be biased toward a certain student demographic group

*Possible miskey because the key should have a positive item-total correlation
**If a multi-point item has less than 1% for a score point or more than 50% 0s, the item is flagged. The rule is 50%+ 0s for mathematics, ELA, and CSLA and 60%+ for science.
***B DIF indicates moderate DIF, whereas C DIF indicates significant DIF.

Separate data review committees are convened for each content area, including the accommodated CSLA items. Participants are provided item images and metadata, along with the classical and DIF statistics. During the data review meetings, educators are trained to interpret the statistical information and judge the appropriateness of the flagged items. The committee members use the data as a tool to direct them toward potential flaws in an item and discuss whether there are construct-irrelevant reasons for a data flag. A data flag, by itself, is not the sole reason an item is rejected. Committee members are instructed that their final judgments about the appropriateness or fairness of an item for any individual and subgroup encompassed by the data flag should be based on their expertise with their content area and experience as Colorado educators.

Committee members review each item and recommend whether to accept or reject it. An accepted item indicates that the educators, through their varying expertise, determined that there is not a construct-irrelevant reason for the data flag within the item, whereas a rejected item indicates that the educators determined there is a construct-irrelevant reason for the data flag. Construct-irrelevant reasons for data flags could include issues such as language that is above grade-level or content that is biased against a particular group. In contrast, construct-relevant explanations could be difficult content that is part of the standards or distractors that reflect a very common misunderstanding of the concept covered by the item, which would not be a reason to reject the item.

Following the data review meetings, CDE reviews the committees' recommendations and makes final decisions. All accepted items are moved into "Ready for Operational" status. Table 3.4 presents the final results following the data review based on Spring 2024 data (i.e., the number of field tested items that were either accepted, accepted for revision and re-field test, or rejected as a result of the data review).

Table 3.4. Data Review Results

Assessment	#Accepted	#Accepted for Revision and Re-Field Test	#Rejected
Mathematics 3	8	3	0
Mathematics 4	8	0	0
Mathematics 5	10	2	0
Mathematics 6	15	6	0
Mathematics 7	14	4	0
Mathematics 8	16	2	0
ELA 3	4	0	2
ELA 4	21	0	3
ELA 5	26	0	1
ELA 6	22	0	5
ELA 7	21	0	3
ELA 8	20	0	3
CSLA 3	45	0	0
CSLA 4	22	0	2
Science 5	20	1	9
Science 8	31	0	11
Science 11	10	1	10

Chapter 4: Test Construction

The Spring 2024 mathematics, ELA, and science grades 5 and 8 operational test forms were newly developed test forms developed by Pearson. The Spring 2024 CSLA forms were newly constructed through an iterative process between Pearson and Tri-Lin. The Spring 2024 CMAS Science grade 11 test form was a subset of the 2023 core form based on the proportional blueprint reduction to reduce testing burden. Appendix N presents the results of a study conducted to evaluate the potential impact of omitting items on the assessment to reduce testing time that helped inform CDE's decision to reduce the blueprint. Once the test forms were constructed, CDE reviewed the forms, provided feedback, and gave final approval. The following guidelines were used during the Spring 2024 form construction:

- Adherence to the test blueprints and test construction specification targets
    - Exact match to blueprint for subclaims
    - Distribution of cognitive complexity that is within range
    - Percentage of TE items that is within range
- Review of the item statistics and adherence to the statistical criteria in the test construction specifications
    - Evaluation of item means, point biserial correlations, and score point distributions
    - Evaluation of item response theory (IRT) item parameter estimates
    - Evaluation of item fit statistics
    - Mirroring of 2018 test characteristic curves (TCCs) and conditional standard error of measurement (CSEM) curves (mathematics and ELA only), mirroring of the 2023 TCC and CSEM for science and CSLA
    - Minimization of CSEM curves around the cut scores
- Balance in the representation of gender, ethnicity, geographic regions, and relevant demographic factors
- Thorough review of individual items to establish that the content within items is up-to-date and relevant
- Selection of items with various stimulus types throughout the test form to enhance the test-taking experience by providing variation in the appearance of item types presented
- Efficient and deliberate use of varied content representative of the knowledge and skills in the ESs or EOs
- Review of the full form, including field test items, for clueing and/or content overlap

4.1. Operational Form Construction

Most students take the CMAS assessments online, which allows for the use of innovative item types and for accessibility features such as text-to-speech and color contrast to be available to all students in both English and Spanish for mathematics and science and in English for the online ELA forms. When building the test forms, assessment specialists select a set of operational items in accordance with the test blueprint and test construction specifications. Items selected for operational use must meet the blueprint requirements and should include a variety of topics and contexts with specified psychometric targets.

For ELA and science, the initial operational item pull is selected first. The assessment specialists verify that the test form meets the blueprint and test construction specifications (i.e., the required ES or EO coverage, claim and subclaim coverage, cognitive complexity allocation, and task type). The form is then presented to a Pearson psychometrician who verifies that the form falls within the established psychometric and blueprint parameters and identifies the anchor item set within each operational form. For mathematics, the initial operational item pull does not begin until after psychometrics has identified the anchor item set. (See Chapter 10 for details about the anchor sets.)

Once the form is vetted internally, the form is presented to CDE for review. If needed, the assessment specialists, Pearson psychometricians, and CDE collaborate to finalize the form. This can be an iterative process, with the result being CDE's approval of the form.

4.2. Field Test Item Selection

After the operational form is approved, field test items are selected from the item bank. The purpose of field testing is to administer newly developed items to generate item statistics and assess their eligibility to become operational items. Items chosen for field testing are placed on a form in a designated section and sequence. Pearson and Tri-Lin assessment specialists assemble field test sets of items so that they comprise the appropriate distribution of standards, subclaims, task types, topic coverage, cognitive levels, and key distributions to meet the required item refresh rates in following years.

4.3. Accommodated Test Forms

Accommodated test forms are available for students who need them and include paper, large print, and braille forms, as well as auditory/signed presentation scripts and online forms designed to work with assistive technology such as screen readers. Auditory/signed presentation scripts are available for the paper forms in both English and Spanish for mathematics and science. English auditory/signed presentation scripts are available for both online and paper forms for local translation into languages other than Spanish including sign language. Due to the effort involved in creating an approved accommodated form, these forms are not refreshed at the same rate as the online forms.

Paper-based test forms are available as an accommodation or for schools that choose not to test online as allowed by state law. CSLA is the accommodated version of CMAS ELA for eligible Spanish-speaking students in grades 3 and 4 and is administered on paper. A Spanish transadaptation paper form is also available for mathematics and science. The paper test form is parallel to the online form (i.e., it includes the same operational items). To support this, parallel paper-based items were developed for TE items in a way that was comparable in terms of student interaction. This was achieved with traditional SR items or required an item that had to be human-scored. For example, a drag-and-drop TE item may have been converted to an item in which the student had to draw lines from the draggers to the drop bays. During equating, the TE item statistics are compared to the paper-based version to confirm equivalence. CSLA also offers paper-based versions of TE items developed to be similar to the ELA paper-based TE items, although all CSLA paper-based TE items are machine scored.

After approval of the paper test materials, a braille version of the assessments is created according to the process outlined below:

1. Pearson Braille Services uses constructed test forms to review the items and clusters for identifying potential modifications related to spacing constraints, visual bias in response expectations, and illustration complexity. Recommendations are documented for modifications to text and images.
2. The modifications document is provided to Pearson assessment specialists to ensure compliance with item constructs and assessed standards.
3. Pearson assessment specialists and CDE review the recommendations and provide feedback regarding any modification concerns.
4. Pearson Braille Services translates the test form into braille and designs print images as tactile graphics.
5. The braille form is proofread by a two-person proof team consisting of a native braille reader, certified as a braille proofreader by the National Library Service, and a sighted copyholder.
6. Edits to text and graphics are made based on the proof team's feedback.
7. The braille form is reviewed by a committee of Pearson staff, CDE staff, and Colorado Teachers of the Visually Impaired (TVI).
8. The braille form is finalized, and hardcopy test books are produced.

Large print versions of the CMAS assessments are a 50% enlargement of the regular paper form and are printed on 14" × 18" paper. When needed, the large print version includes a visual description booklet that contains a description of artwork (maps, photographs) for which it may be difficult for a student with visual impairments to see the subtleties within the art. CDE reviews the paper form and identifies which pieces of art need to be described in the visual description test booklet.

Chapter 5: Test Administration

The CMAS assessments are administered in TestNav, Pearson's online testing platform. PearsonAccess^{next} is the student test management portal Assessment Coordinators and Test Administrators use to manage student tests and registrations and order materials if needed. Prior to the administration of the assessments, districts, schools, and teachers are to ensure that their students and systems are prepared for the assessments. Such information is communicated to the appropriate individuals via manuals, virtual trainings, and recorded modules. Table 5.1 presents the test administration window, including the release of the CMAS score reports. (See Chapter 7 for information on reporting).

Table 5.1. Test Administration Activities

Event	Date(s)
DAC Administration Training	October and November 2023
Spring 2024 Administration Window	April 8–26, 2024
CMAS Student Data Files Available	June 17, 2024
CMAS Reports Available	July 3, 2024

5.1. Manuals

The following manuals were available online at https://coassessments.com/manuals/ to support the CMAS administration:

- The *CMAS Test Administrator Manual* for both online and paper-based testing describes the procedures Test Administrators are to follow when administering the assessments. Test administration policies and procedures are to be followed as written so all testing conditions are uniform statewide. The guidelines and test administration scripts in these manuals are provided to ensure that every student in Colorado receives the same standard directions during the test administration by content area, grade level, and accommodation. Districts receive printed copies of the *CMAS Test Administrator Manual* with their shipment of other CMAS testing materials.
- The *CMAS and CoAlt Procedures Manual* provides instructions for coordination of the CMAS assessments. Instructions include the protocols all school staff are to follow related to test security, test administration, and providing accommodations to students with disabilities and ML students and accessibility features to all students. The manual also includes the tasks to be completed by District Assessment Coordinators (DACs), School Assessment Coordinators (SACs), and District Technology Coordinators (DTCs) before, during, and after the test administration.
- The *PearsonAccess^{next} Online User Guide* provides guidance for DACs, SACs, DTCs, Test Administrators, and student enrollment/sensitive data personnel who use PearsonAccess^{next}.

5.2. Administration Training

Administration training is intended to make sure all individuals involved in CMAS assessment activities at the school and district levels are prepared to follow administration processes and procedures with fidelity, as well as to support adherence to security procedures. Fidelity to standardized test administration processes and procedures helps ensure the comparability of resulting scores and accurate interpretation of results.

Live virtual trainings were conducted by CDE for groups of DACs, during which the DACs independently accessed CDE- and Pearson-developed lessons through an interactive training platform. The lessons contained information regarding proper procedures for administration, security requirements, receiving and returning materials to Pearson, and the use of PearsonAccess^{next} with TestNav. Upon completion of each training lesson, CDE provided additional details pertaining to the covered information and an opportunity for questions and answers. After CDE trained the DACs, the DACs trained the SACs, Test Administrators, and any other individuals within the district who planned to participate in the CMAS administration.

Pearson customer service center staff were also trained to answer questions about the administration and to escalate inquiries as necessary. A knowledge base of common questions was created by CDE and Pearson based on information covered in the training materials and manuals to ensure accurate and consistent responses to school and district personnel, with revisions and additions made as needed. CDE met with Pearson daily during the administration window to review questions from districts and ensure that appropriate answers were provided. Policy questions received by the Pearson customer service center were referred to CDE.

Live webinar accommodations and accessibility features training was also conducted by CDE for district-level personnel to ensure that all individuals providing these supports across the state follow the procedures associated with each accommodation and accessibility feature. Providing accessibility features and accommodations in a standardized manner helps to ensure the comparability of resulting scores and accurate interpretation of results.

Resources used during the live trainings were posted on the CDE Assessment Training website at https://www.cde.state.co.us/assessment/trainings throughout the administration year. Administration training materials such as slide decks, manuals, and how-to guides were also available on this website for training SACs and Test Administrators.

5.3. Practice Resources

Colorado Practice Resources (CPRs) are available online at https://coassessments.com/practice-resources/ to help students become familiar with the CMAS item types. The CPRs are updated as needed to reflect current accessibility features and any updates to TestNav that may impact student interactions with the assessment. Accommodated versions of the CPRs are also available so students can practice using accommodations and accessibility features such as English text-to-speech, color contrast, and Spanish text-to-speech. Paper sample items for students taking the paper versions of the assessments are available in PDF format for download. CPRs are accompanied by scoring guides that include performance metrics and alignment to the CAS.

5.4. Onsite Preparation

Districts were instructed in site readiness preparations, TestNav, proctor caching, and use of the SystemCheck tool to configure their testing technology environments and evaluate their configuration for district readiness. Districts were also provided tools and resources to test their environment readiness status and infrastructure systems.

5.5. Accessibility Features and Accommodations

Accessibility is considered from the beginning of the test development process and is inherent within the CMAS assessment and administration. For example, TestNav includes tools and accessibility features that are available to all students to increase the accessibility of the assessments (e.g., highlighter, online color contrast). Also included is the text-to-speech accessibility feature for mathematics and science that allows for text to be read to students by the embedded software audio feature. Although this feature is available to all students, it is assigned in advance of testing only to those who need it. Similarly, the CSLA assessments were developed as linguistically accommodated Spanish tests to ensure accessibility for eligible Spanish-speaking students.

Accommodations are also available to the population of students with IEP or 504 plans or ML students. For example, students may have extended time as required by their IEP or as allowed for students classified as ML. The test is also available with Spanish text-to-speech (mathematics and science only) and paper transadaptations or auditory presentation scripts that can be translated into other languages. Accommodations are intended to provide equitable access to the assessment without impacting the measured construct. Accommodations can be adjustments to the test presentation, materials, environment, or response mode of the student and are based on individual student need. They should not provide an unfair advantage to any student. Providing an accommodation for the sole purpose of increasing test scores is not ethical.

Accommodations must be documented and used regularly during classroom instruction and assessments prior to the testing window to ensure that the student can successfully use the accommodation. However, although accommodations are used for classroom instruction and assessments, some may not be appropriate for use on statewide assessments. As a result, it is important that educators become familiar with the state assessment policies about the appropriate use of accommodations and that districts have a plan in place to ensure and monitor the appropriate use of accommodations.

Certain accommodations are allowed only in special cases with CDE approval due to being an inherent violation of the intended construct. For example, the accommodations of calculator on non-calculator sections of mathematics and a scribe for CR items for ELA/CSLA require approval to preserve the intended constructs of mathematics and writing according to the CAS.

Some of the available accommodations for CMAS include CSLA in place of ELA (other linguistic accommodations do not apply as CSLA is the linguistic accommodation), English auditory/signed presentation scripts (mathematics and science), Spanish auditory/signed presentation scripts (mathematics and science), auditory/signed presentation scripts for signed presentation and local translation into languages other than English and Spanish, braille forms, large print forms, assistive technology forms for screen readers (mathematics and ELA only), and Spanish forms with and without text-to-speech for mathematics and science.

5.6. Test Security

Test security procedures are put in place to enhance the likelihood that security is maintained before, during, and after the assessment administration. For example, materials used during the paper administration of the assessment are to be kept in locked storage locations when not under the direct supervision of Pearson or approved testing coordinators and administrators. All district and school personnel involved in the CMAS test administration are required to participate in annual local training. DACs are responsible for overseeing training for the district, including verifying that the DTC and SACs are trained. SACs are responsible for ensuring that Test Administrators and all other individuals involved in test administration at the school level are trained and subsequently act in accordance with all security requirements.

A chain of custody plan for materials is required to be written and implemented to ensure that materials are securely distributed from DACs to SACs to Test Administrators and securely returned from Test Administrators to SACs and then to DACs. SACs are required to distribute materials to and collect materials from the Test Administrators each day of testing and to securely store and deliver materials to DACs after testing is completed in accordance with the instructions in the *CMAS and CoAlt Procedures Manual*.

All individuals involved in the test administration are required to sign a security agreement prior to handling test materials, which requires them to follow all procedures set forth in the aforementioned manuals and prevents them from divulging the contents of the assessment, copying any part of the assessment, reviewing test items with the students, allowing students to remove test materials from the testing room, or interfering with the independent work of any student taking the assessment. During online testing, all unnecessary computer functions are disabled, and access is restricted to disallow activities in all applications outside the testing program.

PearsonAccess^{next}, the assessment management system used during the administration, includes permissions-based user role access to all information within the system, including accessing student information, setting up and delivering test sessions (preparing, starting, and stopping sessions), administering tests (unlocking, resuming, and locking units), and accessing reports. Access to the online assessments through the student testing system, TestNav, is tightly controlled before, during, and after test administration, requiring a login ID and password to enter the system for each unit. Test content is locked and cannot be accessed by students or district/school-level users after the students submit their answers. Each unit of the paper test requires students to break the unit seal before accessing the test content. To enhance security during test administration, test forms are spiraled, decreasing the likelihood that a student would be working on the same items as their peers at the same time.

After all test sessions are completed at a school, used and unused materials are required to be securely stored and returned to the DAC by the district deadline for shipment to Pearson. DACs are required to report any missing test materials or test irregularities and to complete the appropriate documentation.

5.7. Assessment Administration Monitoring

Trained test monitors observed live administration of the CMAS assessments in several districts across the state (four to five districts in each of Colorado's eight regions). The monitoring activity is required by the US Department of Education to ensure standardized administration across districts and schools. Colorado statute also requires that CDE review and update administration and security policies as necessary to maintain the integrity of the assessments. CDE gathers information to meet both requirements through this monitoring activity.

Prior to the opening of the CMAS window, DACs were notified by CDE regarding their districts' selection for CMAS monitoring. Districts were selected based on test format (i.e., paper, online, or both), district size, student population, reports of CMAS misadministrations in previous years, and if or when they last had a CMAS test monitor.

Before visiting district schools, test monitors completed background checks, were trained on CMAS administration and the monitoring activity, and contacted DACs to schedule their visits. DACs determined whether they informed schools within their districts regarding the monitoring activity. Upon arrival at the district, test monitors had identification, a letter from CDE stating their purpose at the school, and materials necessary to complete the monitoring activity (e.g., a checklist form for their observations and CMAS directions from the *CMAS Test Administrator Manual* to follow along with administration).

Test monitors met with the DAC for each district, and in some cases the SAC for each school, to go through pre-observation questions about local policies and procedures, then entered the student testing space before the start of the testing session and remained in the space until the session ended. Test monitors completed observation checklists regarding test environment setup, accommodations, Test Administrator adherence to the directions in the Test Administrator Manual, Test Administrator interactions with students, technology interruptions (online testing only), and student testing times. Monitors did not interact with Test Administrators or students.

Observation checklists were submitted and any testing anomalies were reported to the vendor and CDE upon completion of the monitoring activity. CDE staff contacted DACs if issues were observed during test monitor visits to provide guidance and address practices for future administrations. Local feedback from assessment coordinators and any observations indicating the need for updates to policies and procedures were used to make adjustments for future CMAS administrations.

Chapter 6: Scoring

The CMAS assessments use a combination of machine, human, and automated scoring. All SR and online TE items are machine-scored, with point values varying by item type and assessment. Most mathematics and all science CR items are handscored, with a small number of CR mathematics items scored by the Math Reasoning Engine (MRE). The ELA PCR items are scored on two trait dimensions using human scoring, Written Expression (WE) and Knowledge and Use of Language Conventions (WKL). Pearson's Scoring Services team conducted the handscoring for the CR, PCR, and parallel paper-based versions of the TE items for CMAS. To maintain comparability, scoring rules for the machine-scored items and rubrics, anchor papers, rules and scoring methods for the handscored items were preserved from previous years.

6.1. Machine Scoring

Machine-scored items include key-based and rule-based items. Key-based items tend to be a version of multiple-choice and multiple-select (i.e., students select more than one correct answer) items. Rule-based items are machine-scored TE items. Initial scoring expectations are developed during item development and are included in the item review process. The scoring rules and correct responses are included in the items' XML coding. Prior to scoring, key checks and adjudication are completed for all machine-scored items to verify that the machine is correctly identifying correct and incorrect responses. If there is a discrepancy in the scoring, content experts review the item and adjustments are made as needed. During testing, actual distribution of scores is compared to expected distribution. Further evaluation is completed if a discrepancy is identified.

6.2. Handscoring

6.2.1. Operational Scoring

Human-scored operational items are scored using either a distributed or synchronous scoring model depending on the content area. Items on the CSLA form and paper-based TE items are scored synchronously, while scoring for all other human-scored items is completed through distributed scoring. At times, distributed scorers are leveraged to score paper-based TE items. Scoring includes several components that together provide a comprehensive performance scoring model. For example:

- All scorers are required to pass a background check and sign a nondisclosure agreement, agreeing to adhere to all security and confidentiality requirements.
- All scorers have a four-year degree at a minimum. Scorers are assigned to content areas based on their educational backgrounds, related fields of work, and their demonstrated knowledge in the content area.
- Scorers of CSLA items must be proficient in written Spanish and English languages.
- Scorers are trained using comprehensive training materials developed by scoring experts that rely on student responses scored at the rangefinding meetings. Prior to qualifying for an item, scorers review an online training module that includes an overview of scoring; information specific to the item such as the prompt and rubric; and anchor sets. Scorers then score multiple practice sets prior to attempting qualification. After successful qualification, scorers begin scoring the item.

- For CSLA items, training is led by a Pearson scoring director who presents item-specific materials, including the prompt and rubric. The scoring team then receives training on anchor sets prior to moving into the online portion of training where scorers apply scores on multiple practice sets within the electronic scoring system. After each practice set, the scoring director reviews the practice set results with the scorers prior to scorers taking the qualification sets. After successful qualification, scorers begin scoring the item.
- Scorers must pass a qualifying test for the item types that they score. Qualification sets are designed to test scorer accuracy across the range of score points for a given item.
- Student responses are converted to electronic images at Pearson facilities and are then transmitted for computer-based scoring.
- Distributed scorers are located across the United States and work from their homes. Their computers are set up for image-based scoring. A comprehensive set of scoring and monitoring tools are integrated into the scoring system, and content supervisory staff are available by phone to help answer any training or scoring questions. This distributed setup allows scorers to work seven days a week with extended evening hours.
- Synchronous scorers are located across the United States and also work from their homes; however, they are only permitted to score while attending daily Microsoft Teams meetings with content supervisory staff. As with distributed scoring, synchronous scoring uses a comprehensive set of scoring and monitoring tools integrated into the scoring system, with content supervisory staff available within the Microsoft Teams interface to help answer any training or scoring questions. Unlike distributed scoring, synchronous scoring is typically conducted Monday through Friday during normal business hours. Synchronous scorers are used for CSLA forms and paper-based TE items.
- Additional security procedures are in place for distributed scoring. Data are securely transmitted through HTTPS and SSL technology using secure protocols for system authentication. Student responses are randomly routed through the scoring platform to prevent scorer knowledge of student information, unless a student self-identified in the response. Scorers agree not to use shared, institutional, or public computers to score and not to save student responses or test materials. Scorer printing capabilities of materials, such as anchor papers, are only approved for printing after they have undergone and passed a personally identifiable information review by CDE. Scorers agree to securely destroy or return printed materials to Pearson at the conclusion of scoring.

Pearson's processes and tools provide a replicable quality system that strengthens consistency across projects and locations within Pearson's Scoring Services operations. Pearson's Scoring Services team uses a comprehensive system for continually monitoring and maintaining the accuracy of scoring at both the group and individual levels. This system includes daily analysis of a comprehensive set of statistical monitoring reports, as well as regular "backreading" of scorers. Reliability statistics are monitored during scoring, and interventions are applied if a scorer or item is not meeting the minimum requirements.

6.2.2. Field Test Scoring

Embedded field test scoring was completed using synchronous scoring that took place within daily Microsoft Teams meetings. All scorers are required to have a four-year college degree. Field test scorers receive stand-up training led by a Pearson scoring director who presents item-specific materials, including the prompt and rubric. Scorers then review the anchor sets in a group setting prior to scoring practice sets on paper.

6.2.3. Rangefinding

Scoring rubrics are generated for each unique item for mathematics and science, while ELA/CSLA use holistic rubrics for each item type, provided online at https://www.cde.state.co.us/assessment/cmas_testdesign. Rubrics are finalized during rangefinding and are maintained, along with the training materials for each item, by Pearson's Scoring Services group.

Rangefinding meetings take place following the administration in which an item was field tested. As such, rangefinding took place from June 3–6, 2024, following the Spring 2024 administration. The purpose of rangefinding is to define the range of performance levels within the score points of the rubrics using student responses. Each rangefinding committee includes Pearson's Scoring Services and content staff, state content representatives, and educators with relevant grade-level and content expertise and experience with special populations. Participants create consensus scores for a sample set of student responses that are subsequently used to develop effective training materials for scoring of the CR items.

Pearson's scoring directors construct one rangefinding set per item, which includes approximately 30 responses. For multi-point items, pre-constructed sets with additional responses are brought to the meeting. Responses included in these sets represent the full spectrum of scores to the greatest extent possible. The responses for each item are randomly ordered to provide committee members an opportunity to determine the spectrum of scores without bias, although actual scores are not revealed to committee members. Each set includes responses clearly earning each available score point for each item type. The set also includes sample responses that may have been challenging to score (i.e., the score points earned were not necessarily clear).

Following an introductory session presented by a member of the Scoring Services group, the rangefinding committee is divided into several breakout groups based on educator expertise. Each group is assigned a range of field test items to be reviewed based on the following process:

1. The scoring director introduces each item. The committee reviews the item and corresponding rubric.
2. The committee reads student responses—individually or as a group—and then discusses and decides the most appropriate score for each response.
3. The scoring director records committee members' comments and the final consensus score for each student response. Consensus is reached when a majority of committee members agree on a particular score point for a response and all members agree to accept the score of the majority.
4. A designated committee member records consensus scores. After reviewing responses for each item, the committee member compares their notes with those kept by the scoring director and provides sign-off to indicate agreement with the recorded scores.

Following the rangefinding meetings, Scoring Services creates training materials with an anchor set that is used for initial training (up to 15 responses) and a full practice set (up to 10 responses). For ELA, two anchor sets are used per item, one for content and one for conventions. Each CR item is then scored with the associated training materials.

6.2.4. Backreading

Backreading is the method of immediately monitoring a scorer's performance and is an important tool for Pearson's scoring supervisors. Backreading is performed in conjunction with the statistics provided by reader performance reports and as indicated by scoring directors, allowing scoring supervisors to target particular readers and areas of concern. Scorers showing low inter-rater agreement or those showing anomalous frequency distributions are given immediate, constructive feedback and monitored closely until sufficient improvement is demonstrated. Scorers who demonstrate through their agreement rates and frequency distributions that they are scoring accurately continue to be spot-checked as an added confirmation of their accuracy. The agreement rate requirements are as follows. (Refer to Section 11.5 for the inter-rater reliability results.)

- 1-point item: 90% perfect and 95% perfect plus adjacent agreement
- 2-point item: 90% perfect and 95% perfect plus adjacent agreement
- 3-point item: 80% perfect and 95% perfect plus adjacent agreement
- 4-point item: 70% perfect and 95% perfect plus adjacent agreement
- 5+-point item: 65% perfect and 95% perfect plus adjacent agreement

6.2.5. Scoring Calibration Sets

Calibration sets are responses selected as examples that help clarify scoring issues, define more clearly the lines between certain score points, and reinforce the scoring guidelines as presented in the original training sets. They can be applied to groups, a subset of groups, or individual scorers as needed. These sets are used to proactively promote accuracy by exploring project-specific issues, score boundaries, or types of responses that are particularly challenging to score consistently. Scoring directors administer calibration sets as needed, particularly for more difficult items.

6.2.6. Validity Papers

As a quality monitoring tool used during scoring, validity papers are student responses chosen by Pearson scoring directors to measure the accuracy of a scorer when applying the scoring rubric. Validity papers are blind to scorers, which means a scorer is not aware when they are scoring a validity paper. Scoring directors may choose to include an annotation with a validity paper so that a scorer will receive immediate feedback if a validity paper is scored incorrectly. Validity statistics are monitored by scoring directors throughout the life of a scoring project.

6.3. MRE Scoring

The Math Reasoning Engine (MRE) evaluates responses based on rubric criteria specific to the expected item responses and does not require human scoring to train the engine. Because the engine can reason about the student's math, there is no need to enumerate all possible correct responses (as one would with other more traditional machine scoring). The MRE scoring criteria are defined in terms of the mathematical characteristics of a response the engine uses to evaluate the extent to which a student response satisfies a scoring rubric (e.g., equivalence, form, precision, and constraints).

Before a mathematics item is scored operationally, it goes through a rubric refinement and validation process. After field testing, online student responses are aggregated into a set of individual unique responses that are back-read by human scorers. When the human scorers disagree with the score assigned by the scoring engine, they flag the response for further review. The flagged responses are then reviewed by Pearson content experts to determine if scoring rules should be adjusted to better fit the rubric.

If scoring rule adjustments are made, MRE can rescore the entire item based on the new scoring rules configuration. During rescores, every changed response score is identified and reviewed by content experts to evaluate the impact of the rule change on all responses to ensure that the change had the intended effect and nothing more. The data provided to committee members is based on the final scoring rules reflecting how the responses are scored operationally and includes the most common field test item responses providing committee members with additional information to evaluate item quality.

Chapter 7: Reporting

7.1. Available Reports

Two types of score reports are provided: (a) the student-level Student Performance Report and (b) the aggregate reports at the school and district levels. Appendix C presents sample Student Performance Reports, and examples of each type of aggregate report are provided in the *CMAS and CoAlt Interpretive Guide to Assessment Reports*. For a detailed explanation of the information provided in all reports, refer to the *CMAS and CoAlt Interpretive Guide to Assessment Reports* located online at https://www.cde.state.co.us/assessment/cmas_coalt_interpretiveguide_2024. CSLA assessments are parallel and comparable to the CMAS ELA assessments in scoring and reporting. Therefore, separate CSLA reports are not included (please refer to the CMAS ELA examples).

The Student Performance Report provides information about the performance of a particular student. The student's scale score, performance level, percentile ranking, and percent of points possible scores are displayed on a two-page report, along with comparative information related to the student's school, district, and state performance. PLDs are also provided. In addition to the electronic versions made available to districts and schools, two copies of the Student Performance Report are printed and shipped to districts for distributing to parents/guardians and for maintaining locally.

The following aggregate reports are produced at the school and/or district levels and provide summary information for a given school or district. They are provided electronically through PearsonAccess^{Next}, with access limited to authorized users. The participation report provides a comparison of the demographic characteristics of the tested students compared to all students eligible for testing. This information can assist districts and schools in determining how to interpret their aggregated results.

- Performance Level Summaries
- Content Standards Rosters
- Evidence Statement Analysis Reports (mathematics and ELA only)
- Item Analysis Report (science only)
- District Summary of Schools (district level only)
- District and School Participation Reports

7.2. Interpretation of Test Scores

The CMAS reports provide information on student performance in terms of scale scores, performance levels, percentile ranks, and percent earned scores, as described below.

7.2.1. Scale Scores

A scale score is a conversion of a student's response pattern to a common scale that allows for a numerical comparison between students. Scale scores are particularly useful for comparing test scores over time and creating comparable scores when a test has multiple forms. All CMAS assessments provide an overall scale score. ELA reports also include a scale score for the Reading claim, and the science reports provide separate scale scores for content standards and SEPs (referred to as reporting categories).

The overall scale for each content area assessment ranges from 650 to 850, and the ELA Reading scale ranges from 110 to 190. The science content standards scale score ranges from 400 to 600, although the graph displayed on the student reports ends at 550. Any student who earned a score greater than 550 still had their score written on the report but the diamond representing this performance would appear at the end of the graph at 550.

7.2.2. Performance Levels and PLDs

The performance levels are based on the overall scale score, and cut scores divide the score scale for a grade and content area into the performance levels. (Refer to Chapter 8 for information on the cut scores.) The CMAS Mathematics and ELA assessments have five performance levels (*Did Not Yet Meet Expectations*, *Partially Met Expectations*, *Approached Expectations*, *Met Expectations*, and *Exceeded Expectations*), whereas CMAS Science has four performance levels (*Partially Met Expectations*, *Approached Expectations*, *Met Expectations*, and *Exceeded Expectations*). Students in the top two performance levels met or exceeded the expectations of the CAS and are considered on track for the next grade level in that content area.

The performance levels are accompanied by performance level descriptors (PLDs) that articulate what a student should know and be able to do in a particular performance level (e.g., the set of statements describing what it means for a grade 8 student to reach *Met Expectations* in mathematics). The CMAS assessments use two types of PLDs: (a) policy PLDs (also known as policy claims) that provide a general idea of what is expected of a student at each level regardless of their grade level, as shown in Table 7.1 and Table 7.2, and (b) grade-level PLDs that provide detailed descriptions of performance levels by grade level and content area, available online at https://www.cde.state.co.us/assessment/cmas_plds and included on the Individual Student Performance Report and in the *CMAS and CoAlt Interpretive Guide to Assessment Reports*.

Table 7.1. Performance Levels and Policy Claims—Mathematics and ELA

Performance Level	*Did Not Yet Meet Expectations*	*Partially Met Expectations*	*Approached Expectations*	*Met Expectations*	*Exceeded Expectations*
Policy Claim	Students who do not yet meet academic expectations for the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They will *need extensive academic support* to engage successfully in further studies in this content area.	Students who demonstrate a limited command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They will *need additional academic support* to engage successfully in further studies in this content area.	Students who demonstrate a moderate command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They will *likely need additional academic support* to engage successfully in further studies in this content area.	Students who demonstrate a strong command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They are *academically prepared* to engage successfully in further studies in this content area.	Students who demonstrate a distinguished command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They are *academically well prepared* to engage successfully in further studies in this content area.
Scale Score	650–699	700–724	725–749	750–varies*	varies*–850

*Varies by grade and content area

Table 7.2. Performance Levels and Policy Claims—Science

Performance Level	*Partially Met Expectations*	*Approached Expectations*	*Met Expectations*	*Exceeded Expectations*
Policy Claim	Students who demonstrate a limited command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They will *need additional academic support* to engage successfully in further studies in this content area.	Students who demonstrate a moderate command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They will *likely need additional academic support* to engage successfully in further studies in this content area.	Students who demonstrate a strong command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They are *academically prepared* to engage successfully in further studies in this content area.	Students who demonstrate a distinguished command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They are *academically well prepared* to engage successfully in further studies in this content area.
Scale Score	650–724	725–749	750–varies*	varies*–850

*Varies by grade

7.2.3. Percentile Ranking

Percentile rankings are provided on student performance reports to indicate how the student performed compared with other students in the state. For example, a student with a percentile ranking of 70 performed better than 70% of students in Colorado. The percentile rankings are based on the overall scale score.

7.2.4. Percent Earned

To prevent incorrect interpretations and provide a metric that is more generally understood, student performance for subclaims and the Writing claim (ELA/CSLA) are reported as the percentage of points earned (i.e., the number of points a student earned out of the total number of points possible within a claim or subclaim). The overall Writing claim is calculated as the sum of the individual trait scores, Written Expression (WE) and Knowledge of Language and Conventions (WKL), with WE scores multiplied by three to emphasize the additional importance of the content of the writing as reflected in the standards. The percent of points earned for WE is the sum of the unweighted WE trait scores on both PCRs, and WKL is the sum of the WKL trait scores on both PCRs.

Unlike scale scores, the percent of points possible scores cannot be compared across years because individual items change from year to year and are not constructed to be comparable in difficulty at the claim, subclaim, or subscale level. Performance on different subclaims or subscales also cannot be compared within an administration because the number of items and the difficulty of the items within each claim, subclaim or subscale may not be the same.

The percent of points possible can be compared to aggregated state, district, and school performance. The student performance reports also include an indicator of how students who scored just above the *Met Expectations* cut score on the overall assessment performed on each category. This indicator gives similar information to the *Met Expectations* cuts.

Chapter 8: Standard Setting

To support the interpretation of student results, student performance on the CMAS assessments is described in terms of performance levels as presented in Table 7.1. Standard setting is the process of translating those policy-driven performance standards into scores on the assessment. The purpose of standard setting is to determine the boundaries—or cut scores—along the score scale that differentiate student performance among those levels (e.g., Cizek et al., 2004; Kane, 1994).

Table 8.1 presents the cut scores for each content area and grade. The mathematics and ELA cut scores were set in 2015 in collaboration with the PARCC consortium using the Evidence-Based Standard Setting (EBSS) method (Beimers et al., 2012), as detailed in the 2015 PARCC *Performance Level Setting Technical Report* (Davis & Moyer, 2015). CSLA cut scores were set in 2016 using the Modified Extended Angoff method, as detailed in the *CSLA Colorado Spanish Language Arts Standard Setting Report* (CDE, 2016). Standard setting for the new science assessment took place from September 27–28, 2022, using a modified version of the Item Descriptor (ID) Matching method (Ferrara et al., 2008), as detailed in the *CMAS Science 2022 Standard Setting Report* (Pearson, 2024).

Table 8.1. Performance Level Cut Scores

Assessment	*Did Not Yet Meet Expectations*	*Partially Met Expectations*	*Approached Expectations*	*Met Expectations*	*Exceeded Expectations*
Mathematics 3	650–699	700–724	725–749	750–789	790–850
Mathematics 4	650–699	700–724	725–749	750–795	796–850
Mathematics 5	650–699	700–724	725–749	750–789	790–850
Mathematics 6	650–699	700–724	725–749	750–787	788–850
Mathematics 7	650–699	700–724	725–749	750–785	786–850
Mathematics 8	650–699	700–724	725–749	750–800	801–850
ELA 3	650–699	700–724	725–749	750–809	810–850
ELA 4	650–699	700–724	725–749	750–789	790–850
ELA 5	650–699	700–724	725–749	750–798	799–850
ELA 6	650–699	700–724	725–749	750–789	790–850
ELA 7	650–699	700–724	725–749	750–784	785–850
ELA 8	650–699	700–724	725–749	750–793	794–850
CSLA 3	650–699	700–724	725–749	750–778	779–850
CSLA 4	650–699	700–724	725–749	750–771	772–850
Science 5	–	650–724	725–749	750–788	789–850
Science 8	–	650–724	725–749	750–796	797–850
Science 11	–	650–724	725–749	750–786	787–850

The ELA assessment also includes a Reading score that has the same range and cut score for all grades. There is only one cut score that corresponds to the *Met Expectations* overall performance level, as shown in Table 8.2, that was determined using the cut information from setting the standards on the overall ELA test (i.e., it was not set separately at the standard setting meeting).

Table 8.2. ELA Reading *Met Expectations* Cut Score

Scale Range	Cut Score
110–190	150

Science also includes performance indicator cut scores that indicate average performance in each reporting category compared to the state, as shown in Table 8.3. These cuts are not used for accountability and change from year to year with shifts in the average performance. Students with scores below this range scored "lower than average" in the reporting category, and students above the range scored "higher than average."

Table 8.3. 2024 CMAS Science Performance Indicator Cut Scores

Assessment	Physical Science	Life Science	Earth and Space Science	SEPs
Science 5	450–520	446–523	449–521	452–519
Science 8	443–515	441–516	438–516	446–514
Science 11	445–511	440–513	437–512	447–509

Chapter 9: Test Results and Analysis

9.1. Student Participation

Table 9.1 presents the number of students who took the Spring 2024 assessment online and those who took the accommodated forms, and Appendix D presents the n-counts by demographic subgroup. Most students took the assessments online.

Table 9.1. Student Participation by Form

Content Area	Form	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
Mathematics	Online	38,598	40,582	41,451	46,382	46,188	44,153	–
	Spanish Online	1,752	1,453	1,085	586	619	546	–
	Paper	3,338	3,307	2,847	2,436	2,452	2,180	–
	Spanish Paper	154	114	54	9	9	9	–
	Text-to-Speech	12,884	11,936	11,159	5,703	4,629	3,906	–
	Assistive Technology	3	3	5	8	5	9	–
	Total	56,729	57,395	56,601	55,124	53,902	50,803	–
ELA	Online	50,919	52,196	52,737	52,040	50,973	48,257	–
	Paper	3,654	3,508	3,211	2,513	2,347	2,073	–
	Assistive Technology	4	6	5	9	14	11	–
	Total	54,577	55,710	55,953	54,562	53,334	50,341	–
CSLA	Paper	1,560	1,162	–	–	–	–	–
Science	Online	–	–	41,933	–	–	43,789	32,021
	Spanish Online	–	–	371	–	–	281	244
	Paper	–	–	2,142	–	–	1,812	644
	Spanish Paper	–	–	16	–	–	10	29
	Text-to-Speech	–	–	10,996	–	–	3,786	1,110
	Spanish Text-to-Speech	–	–	627	–	–	236	60
	Total	–	–	56,085	–	–	49,914	34,108

9.2. Performance Results

Table 9.2 presents summary statistics for the overall scale scores, including the mean, standard deviation (SD), and median, and Table 9.3 presents the percentage of students classified into each performance level based on their overall scale scores. The previous year's results are also included for comparison. Appendix E presents the cumulative scale score distributions by grade, Appendix F displays the results in graphical form, and Appendix G presents the summary statistics for the overall scale scores by demographic subgroup.

Table 9.2. Scale Score Performance Summary

Assessment	2024 N	2024 Mean	2024 SD	2024 Median	2023 N	2023 Mean	2023 SD	2023 Median
Mathematics 3	56,729	740	37.5	741	57,382	738	39.5	739
Mathematics 4	57,395	735	34.7	733	56,789	733	33.5	733
Mathematics 5	56,601	739	33.7	737	56,896	737	35.2	736
Mathematics 6	55,124	732	31.6	730	55,913	730	33.3	729
Mathematics 7	53,902	733	29.5	732	54,148	731	28.1	729
Mathematics 8	50,803	731	41.3	726	52,036	732	40.9	728

Assessment	2024 N	2024 Mean	2024 SD	2024 Median	2023 N	2023 Mean	2023 SD	2023 Median
ELA 3	54,577	738	44.1	740	55,737	737	43.6	738
ELA 4	55,710	741	37.3	742	55,519	742	36.9	744
ELA 5	55,953	747	33.8	747	56,657	747	33.8	747
ELA 6	54,562	743	33.1	744	55,602	743	33.0	744
ELA 7	53,334	746	38.8	746	53,895	744	38.0	745
ELA 8	50,341	740	41.2	742	51,760	741	40.7	742
ELA Reading 3	54,577	145	17.8	146	55,737	145	17.7	145
ELA Reading 4	55,710	147	15.0	146	55,519	147	14.8	147
ELA Reading 5	55,953	149	13.6	149	56,657	149	13.6	149
ELA Reading 6	54,562	147	13.2	147	55,602	147	13.2	147
ELA Reading 7	53,334	149	15.5	148	53,895	148	15.1	148
ELA Reading 8	50,341	146	16.4	147	51,760	147	16.2	147
CSLA 3	1,560	723	26.6	723	1,440	724	27.3	725
CSLA 4	1,162	720	28.3	722	1,180	725	23.8	725
CSLA Reading 3	1,560	139	9.7	139	1,440	140	10.2	139
CSLA Reading 4	1,162	138	10.9	138	1,180	140	9.3	140
Science 5	56,085	736	33.2	740	56,428	733	33.9	737
Science 8	49,914	731	33.8	735	50,947	731	33.0	735
Science 11	34,108	730	29.0	732	31,767	729	29.5	732

Table 9.3. Performance Level Distribution

Assessment	2024 1	2024 2	2024 3	2024 4	2024 5	2023 1	2023 2	2023 3	2023 4	2023 5
Mathematics 3	15.17	19.31	23.80	32.16	9.56	17.76	18.62	23.25	31.02	9.34
Mathematics 4	17.34	23.89	24.70	29.72	4.36	17.38	23.48	26.46	29.75	2.93
Mathematics 5	12.62	24.24	25.82	30.30	7.02	15.17	23.48	24.85	29.05	7.44
Mathematics 6	15.57	28.28	26.95	24.93	4.28	19.93	24.86	26.97	24.17	4.07
Mathematics 7	11.78	29.20	29.23	25.69	4.10	12.10	31.95	29.69	23.46	2.79
Mathematics 8	24.60	23.84	19.05	26.19	6.33	23.74	23.27	20.32	26.98	5.68
ELA 3	22.01	15.55	20.33	37.74	4.36	22.19	17.03	20.91	34.92	4.95
ELA 4	14.52	18.98	24.51	32.02	9.97	14.52	16.47	25.22	35.05	8.72
ELA 5	8.16	17.85	26.70	40.83	6.45	7.93	18.51	25.80	41.11	6.65
ELA 6	10.16	19.80	26.05	36.17	7.83	10.31	20.16	26.16	35.75	7.62
ELA 7	12.81	18.11	22.75	29.25	17.07	13.23	17.52	24.24	30.35	14.67
ELA 8	18.07	16.90	22.23	33.06	9.74	16.54	17.95	23.14	32.12	10.26
CSLA 3	17.05	33.33	32.37	15.77	1.47	19.93	29.65	31.74	16.81	1.88
CSLA 4	25.22	27.88	30.98	13.68	2.24	13.98	34.07	37.71	12.20	2.03
Science 5	–	33.11	28.96	34.16	3.76	–	35.33	30.79	31.15	2.73
Science 8	–	37.73	30.03	31.72	0.53	–	38.53	30.18	30.76	0.53
Science 11	–	37.79	37.41	23.70	1.10	–	38.79	36.63	23.84	0.75

Note. 1 = *Did Not Yet Meet Expectations*, 2 = *Partially Met Expectations*, 3 = *Approached Expectations*, 4 = *Met Expectations*, 5 = *Exceeded Expectations*. Percentages may not sum to 100 due to rounding.

Appendix H presents the summary statistics for points earned by subclaim. While the overall scale scores and Reading scale scores are comparable to results from previous administrations, the assessments are not designed to permit meaningful comparisons across percent earned scores, either within an assessment or across administration years. The difficulty of the items that make up each subscore can vary across subscores and from year to year, making it inappropriate to make inferences based on percent earned performance across subscores or based on subscore performance across years. The only percent earned subscore comparisons supported by the CMAS assessments are those comparing individual or group performance within one subclaim with the performance of other students or groups within the same subclaim and administration.

9.3. Classical Item Analysis

Table 9.4 and Table 9.5 present the overall item difficulty and item discrimination results across all items for each assessment, and Appendix I presents the item-level classical statistics, including the omit rate, *p*-value, item-total correlation, and the percentage of students earning each score point for the CR items.

Item difficulty is measured by the *p*-value bounded by 0.0 and 1.0. The *p*-value for 1-point items is the proportion of students who answered an item correctly and is calculated by dividing the number of students who got the item correct by the total number of students who answered it. For multiple-point items, the *p*-value is the average item score (i.e., the sum of student scores on an item divided by the total number of students who responded to the item) that is then put on a 0 to 1 scale by dividing the average item score by the maximum number of points for the item. A high *p*-value indicates that an item is easy (high proportion of students answered it correctly), whereas a low *p*-value indicates that an item is difficult. Easy and hard items are both necessary to include on an assessment to balance the test difficulty.

Item discrimination is represented by the item-total correlation (also known as the point-biserial correlation) that is bounded by -1.0 and 1.0 and indicates how well an item discriminates, or distinguishes, between low-performing and high-performing students. The item-total correlation is based on the relationship between student performance on a specific item and performance on the entire test based on their test score. Students who do well on a test are expected to do well on a given item, and students who do not do well on a test are expected to not do well on a given item. This means that for a highly discriminating item, students who get the item correct will have a higher average test score than students who get the item incorrect. An item with a high positive item-total correlation discriminates between low-performing and high-performing students better than an item with an item-total correlation near zero. A negative item-total correlation indicates that low-performing students did better on that item than high-performing students.

Table 9.4. Summary of *P*-Values

Assessment	#OP Items	Mean	SD	Min.	Max.	Median
Mathematics 3	33	0.51	0.19	0.22	0.82	0.49
Mathematics 4	30	0.45	0.17	0.20	0.75	0.45
Mathematics 5	30	0.46	0.16	0.20	0.71	0.47
Mathematics 6	29	0.34	0.14	0.11	0.62	0.32
Mathematics 7	29	0.35	0.14	0.14	0.69	0.31
Mathematics 8	28	0.36	0.15	0.06	078	0.36

Assessment	#OP Items	Mean	SD	Min.	Max.	Median
ELA 3	21	0.52	0.14	0.18	0.72	0.56
ELA 4	23	0.44	0.15	0.20	0.76	0.40
ELA 5	22	0.46	0.15	0.21	0.78	0.44
ELA 6	24	0.48	0.12	0.24	0.71	0.49
ELA 7	25	0.48	0.09	0.30	0.65	0.46
ELA 8	25	0.47	0.14	0.24	0.73	0.47
CSLA 3	21	0.31	0.11	0.09	0.58	0.28
CSLA 4	23	0.35	0.16	0.14	0.63	0.34
Science 5	39	0.46	0.17	0.14	0.87	0.47
Science 8	47	0.38	0.18	0.09	0.75	0.36
Science 11	32	0.39	0.18	0.07	0.80	0.42

Note. SD = standard deviation, Min. = minimum, Max. = maximum

Table 9.5. Summary of Item-Total Correlations

Assessment	#OP Items	Mean	SD	Min.	Max.	Median
Mathematics 3	33	0.55	0.11	0.33	0.79	0.53
Mathematics 4	30	0.56	0.12	0.29	0.79	0.56
Mathematics 5	30	0.57	0.12	0.37	0.80	0.57
Mathematics 6	29	0.55	0.13	0.33	0.79	0.53
Mathematics 7	29	0.54	0.16	0.22	0.84	0.51
Mathematics 8	28	0.54	0.15	0.27	0.81	0.53
ELA 3	21	0.59	0.08	0.46	0.77	0.60
ELA 4	23	0.51	0.12	0.36	0.78	0.49
ELA 5	22	0.52	0.14	0.26	0.81	0.54
ELA 6	24	0.53	0.11	0.38	0.82	0.52
ELA 7	25	0.55	0.12	0.38	0.84	0.54
ELA 8	25	0.52	0.13	0.27	0.86	0.54
CSLA 3	21	0.50	0.14	0.25	0.79	0.52
CSLA 4	23	0.52	0.16	0.16	0.83	0.56
Science 5	39	0.45	0.14	0.16	0.71	0.44
Science 8	47	0.44	0.13	0.17	0.69	0.44
Science 11	32	0.45	0.14	0.18	0.69	0.44

Note. SD = standard deviation, Min. = minimum, Max. = maximum

9.4. Subclaim Correlations

The ELA/CSLA tests include Reading and Writing claim scores and five subclaim scores: Reading: Literary Text (RL), Reading: Informational Text (RI), Reading: Vocabulary (RV), Writing: Written Expression (WE), and Writing: Knowledge and Use of Language Conventions (WKL). The Reading score is a composite of RL, RI, and RV, and the Writing score is a composite of WE and WKL reported as a percentage of points earned. It comprises PCR items only. The operational test analyses were performed by evaluating the separate trait scores of WE and WKL. Some PCR items also include RL or RI points, but the reading points for those items were a duplicate of the WE score and were not included in calibrations.

The mathematics tests have four subclaim scores: Subclaim A: Major Content, Subclaim B: Additional & Supporting Content, Subclaim C: Expressing Mathematical Reasoning, and Subclaim D: Modeling & Application. The science test has four reporting categories: Physical Science, Life Science, Earth Systems Science, and Science and Engineering Practices (SEPs). The SEP score is based on a combination of items that are also included in the content standard scores so it is not included separately in the correlation analysis.

Table 9.6 – Table 9.10 present the correlations among subscores, which is one way to assess the internal structure of a test. The ELA/CSLA analyses were conducted between the Reading and Writing claim scores and the subclaims (RL, RI, RV, WE, and WKL), and the mathematics and science analyses were conducted between the subclaim or content standard scores. The intercorrelations for the mathematics and science subscores were higher overall than the ELA/CSLA intercorrelations. Correlations between subscores for mathematics and science ranged from 0.605 to 0.943, while for ELA/CSLA they ranged from 0.373 to 0.931. For ELA/CSLA, the two writing subclaims tended to have higher correlations with one another than they did with any of the reading subclaims. Correlations between the subclaims and the total test ranged from 0.670 to 0.959. There is evidence of unidimensionality if the components within a content area are strongly related to each other.

Table 9.6. Correlations Between Subclaims—Mathematics

Grade	Subclaim	Subclaim B	Subclaim C	Subclaim D	Total Test
3	A	0.814	0.741	0.754	0.952
	B	–	0.702	0.720	0.892
	C	–	–	0.733	0.867
	D	–	–	–	0.870
4	A	0.750	0.788	0.771	0.959
	B	–	0.697	0.690	0.837
	C	–	–	0.735	0.895
	D	–	–	–	0.870
5	A	0.765	0.755	0.742	0.955
	B	–	0.691	0.702	0.858
	C	–	–	0.735	0.870
	D	–	–	–	0.866
6	A	0.730	0.755	0.752	0.946
	B	–	0.673	0.658	0.841
	C	–	–	0.703	0.877
	D	–	–	–	0.862
7	A	0.665	0.805	0.757	0.950
	B	–	0.640	0.605	0.774
	C	–	–	0.783	0.916
	D	–	–	–	0.873
8	A	0.739	0.807	0.713	0.959
	B	–	0.703	0.629	0.836
	C	–	–	0.724	0.904
	D	–	–	–	0.827

Table 9.7. Correlations Between Subclaims—ELA

Grade	Subclaim	RI	RV	WE	WKL	Total Test
3	RL	0.756	0.703	0.675	0.537	0.912
	RI	–	0.723	0.615	0.531	0.893
	RV	–	–	0.571	0.494	0.842
	WE	–	–	–	0.600	0.820
	WKL	–	–	–	–	0.670
4	RL	0.658	0.659	0.677	0.587	0.888
	RI	–	0.619	0.528	0.506	0.828
	RV	–	–	0.550	0.506	0.783
	WE	–	–	–	0.765	0.851
	WKL	–	–	–	–	0.761
5	RL	0.660	0.687	0.628	0.596	0.871
	RI	–	0.605	0.662	0.617	0.854
	RV	–	–	0.597	0.543	0.797
	WE	–	–	–	0.828	0.875
	WKL	–	–	–	–	0.809
6	RL	0.738	0.639	0.628	0.606	0.866
	RI	–	0.666	0.699	0.668	0.902
	RV	–	–	0.530	0.505	0.756
	WE	–	–	–	0.882	0.881
	WKL	–	–	–	–	0.836
7	RL	0.779	0.727	0.723	0.692	0.909
	RI	–	0.712	0.637	0.630	0.888
	RV	–	–	0.570	0.572	0.809
	WE	–	–	–	0.903	0.878
	WKL	–	–	–	–	0.849
8	RL	0.751	0.688	0.644	0.643	0.857
	RI	–	0.695	0.750	0.732	0.914
	RV	–	–	0.565	0.569	0.782
	WE	–	–	–	0.931	0.906
	WKL	–	–	–	–	0.884

Note. RL = Reading: Literary Text, RI = Reading: Informational Text, RV = Reading: Vocabulary, WE = Writing: Written Expression, WKL = Writing: Knowledge and Use of Language Conventions.

Table 9.8. Correlations Between Subclaims—CSLA

Grade	Subclaim	RI	RV	WE	WKL	Total Test
3	RL	0.601	0.671	0.689	0.542	0.891
	RI	–	0.562	0.475	0.385	0.739
	RV	–	–	0.543	0.461	0.797
	WE	–	–	–	0.666	0.866
	WKL	–	–	–	–	0.709
4	RL	0.632	0.678	0.750	0.614	0.924
	RI	–	0.534	0.510	0.409	0.747
	RV	–	–	0.543	0.373	0.741
	WE	–	–	–	0.718	0.896
	WKL	–	–	–	–	0.728

Note. RL = Reading: Literary Text, RI = Reading: Informational Text, RV = Reading: Vocabulary, WE = Writing: Written Expression, WKL = Writing: Knowledge and Use of Language Conventions.

Table 9.9. Correlations between Claims—Reading vs. Writing

Content Area	Grade	Correlation
ELA	3	0.719
	4	0.688
	5	0.731
	6	0.718
	7	0.721
	8	0.748
CSLA	3	0.685
	4	0.725

Table 9.10. Correlations Between Content Standards—Science

Grade	Content Standard	Life Science	Earth and Space Science	Total Test
5	Physical Science	0.731	0.804	0.938
	Life Science*	–	0.716	0.860
	Earth and Space Science	–	–	0.934
8	Physical Science	0.787	0.755	0.913
	Life Science	–	0.774	0.943
	Earth and Space Science	–	–	0.904
11	Physical Science	0.728	0.726	0.922
	Life Science	–	0.694	0.894
	Earth and Space Science	–	–	0.882

*For grade 5, the content standard is Physical Science/Life Science.

Chapter 10: Calibration, Equating, and Scaling

Item response theory (IRT) was used to develop, calibrate, equate, and scale the CMAS assessments. All test analyses including calibration, scaling, and item model fit were accomplished within the IRT framework. The CMAS Mathematics and ELA scales were equated to the previous CMAS (i.e., PARCC) base scale. The calibration of the first operational administration determined the base scale for CSLA and CMAS Science.

Calibration is the process of estimating the parameters (such as item difficulty) for each item on an assessment so that all items are placed on a common scale. To maintain the same performance standards across different administrations of a particular test, it is necessary for each administration of the test to be of comparable difficulty. It is not fair to compare students to a common standard if the overall difficulty of the forms changes from year to year. Maintaining test form difficulty across administrations is achieved through equating. Equating adjusts for differences in overall test difficulty of test forms so that the scores resulting from two different administrations can be considered interchangeable. Equating and scaling typically occur in sequence. First, equating is used to adjust for differences in test difficulty so resulting estimates of student proficiency (i.e., equated raw scores, theta estimates) are on a common metric. The equated estimates of proficiency are then converted to scale scores for reporting purposes.

Table 10.1 summarizes the Spring 2024 calibration, equating, and scaling processes for the operational and field test items. All assessments were post-equated. The entire process was completed for each grade-level assessment, and all steps were independently replicated by at least two members of the Pearson psychometrics team to ensure accuracy.

Table 10.1. Summary of Calibration, Equating, and Scaling Processes

Assessment	Operational Items	Embedded Field Test Items
CMAS	• Obtain the online operational item parameter estimates using IRTPRO control files and IDM. • Evaluate the consistency of scoring and stability of the anchor items. • Scale the operational items to the operational scale using STUIRT. • Calculate item fit statistics and plot expected vs. observed IRFs for each operational item. • Estimate student abilities using IRT score estimation (ISE).	• Obtain item parameter estimates of the operational and field test items using IRTPRO control files and IDM. • Scale the field test items to the operational scale using STUIRT and the online operational items as the anchor set. • Calculate item fit statistics and plot expected vs. observed IRFs for each field test item.
CSLA	• Obtain the non-anchor operational item parameter estimates by fixing the anchor items' item parameter estimates using Winsteps control files and IDM. • Evaluate the stability of the anchor items to obtain the final anchor set. • Scale the non-anchor items to the operational scale using the final anchor set in Winsteps • Obtain item difficulty values, step deviation values, and item fit values for all items. • Estimate student abilities using Winsteps.	• Scale the field test item parameter estimates to the operational scale by fixing the item parameter estimates of the operational items using Winsteps control files and IDM. • Obtain field test item difficulty values, step deviation values, and item fit values for each field test item.

Note. IRF = item response function, IDM = incomplete data matrix. Sources: IRTPRO (Vector Psychometric Group, 2022), STUIRT (Kim & Kolen, 2004), ISE (Chien & Shin, 2012), Winsteps (Linacre, 2011)

10.1. IRT Models

The two-parameter logistic (2PL; Birnbaum, 1968) and generalized partial credit (GPC; Muraki, 1992) models were applied to CMAS Mathematics and ELA; the 2PL, three-parameter logistic (3PL; Birnbaum, 1968), and GPC models were applied to CMAS Science; and the Rasch partial credit model (RPCM) was applied to CSLA. The 2PL model uses two item parameters to relate the probability of person *i* correctly answering a dichotomously scored item *j*:

$$P_{ij}(\theta) = \frac{1}{1 + \exp\left[-Da_j\left(\theta_i - b_j\right)\right]}$$

where *D* is set equal to 1 when defined on the logistic scale, as IRTPRO parameterizes all models. The item discrimination parameter is a_j, and the item difficulty parameter is b_j. The 3PL model adds an item parameter:

$$P_{ij}(\theta) = c_j + \frac{1 - c_j}{1 + \exp\left[-Da_j\left(\theta_i - b_j\right)\right]}$$

where c_j is the item pseudo-guessing parameter. The GPC model has three item parameters to relate the probability of person *i* responding in the *x*-th category to a polytomous scored item *j*:

$$P_{ij}(\theta) = \frac{\exp\left[\sum_{v=0}^{x} Da_j\left(\theta_i - b_j + d_{jv}\right)\right]}{\sum_{k=0}^{M_i} \exp\left[\sum_{v=0}^{k} Da_j\left(\theta_i - b_j + d_{jv}\right)\right]}, x = 0, 1, \ldots, M_j$$

where all parameters are as they were before, and d_{jv} is the category parameter for category *v* of item *j* and M_j is the maximum score on item *j.* To put the parameters on the normal ogive metric, the a_j is then divided by 1.7.

RPCM used for CSLA is an extension of the Rasch one-parameter IRT model attributed to Georg Rasch (1966), as extended by Wright and Stone (1979), Masters (1982), and Wright and Masters (1982). RPCM is a mathematical measurement model with a single item parameter relating a student's performance on a given item involving *m*+1 score categories. The probability of student *n* scoring *x* on *m* steps of item *i* is a function of the student's performance level, θ_n (also referred to as "ability"), and the step difficulties, δ_{ij} of the *m* steps in question *i* as follows:

$$P_{xni} = \frac{exp\sum_{j=0}^{x}\left(\theta_n - \delta_{ij}\right)}{\sum_{k=0}^{m_i} exp\sum_{j=0}^{k}\left(\theta_n - \delta_{ij}\right)}, x = 0, 1, \ldots m_i$$

10.2. Item Response and Characteristic Curves

The item response functions (IRFs) of the IRT models described above relate student ability to the probability of observing a particular item response given the item's characteristics, whereas the item characteristic function (ICF) relates student ability to the expected student score. The graphical representation of the IRF and ICF are the item response curves (IRCs) and item characteristic curves (ICCs), respectively. The IRF and ICF are the same for dichotomous items but differ for polytomous items.

Consider Figure 10.1 that depicts a 2PL item that falls at approximately 0.85 on the ability scale. When a student answers an item at the same level as their ability, they have a roughly 50% probability of answering the item correctly (i.e., in a group of 100 students who all have an ability of 0.85, about 50% of them would be expected to answer the item correctly). A student whose ability is above 0.85 would have a higher probability of getting the item right, while a student whose ability is below 0.85 would have a lower probability of getting the item right.

Figure 10.1. Example 2PL ICC

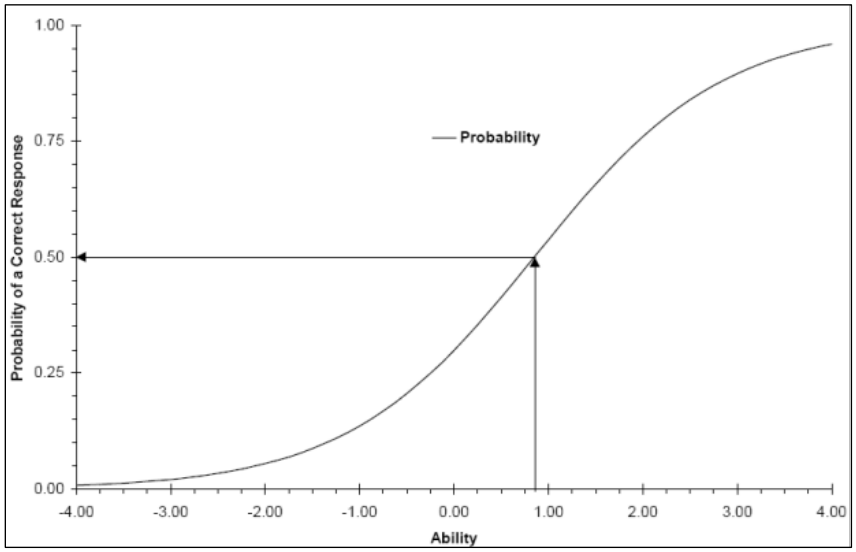


Figure 10.2 presents IRCs of obtaining a wrong answer or a right answer. The dotted line (j=0) shows the probability of getting a score of 0, while the solid line (j=1) shows the probability of getting a score of 1. The point at which the two curves cross indicates the transition point on the ability scale where the most likely response changes from a 0 to a 1. At this intersection, the probability of answering the item correctly is 50%.

Figure 10.2. Example 2PL IRC

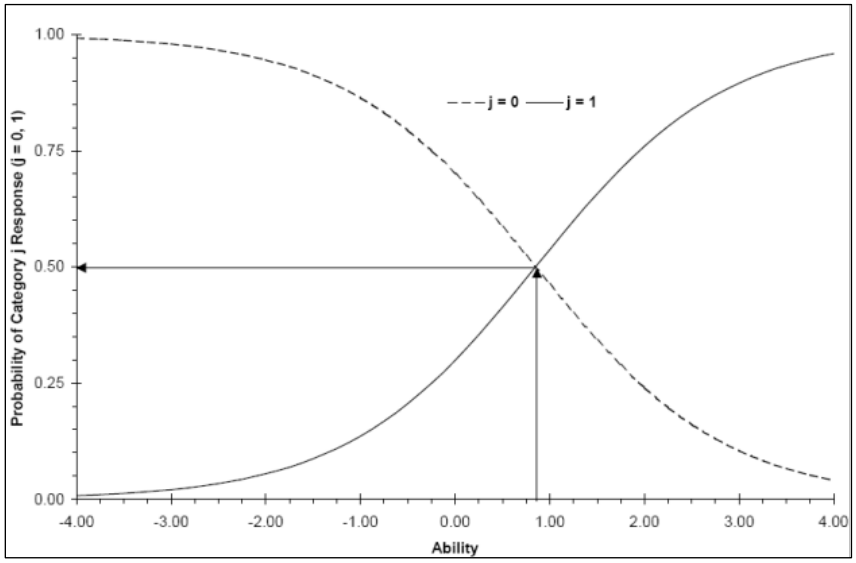
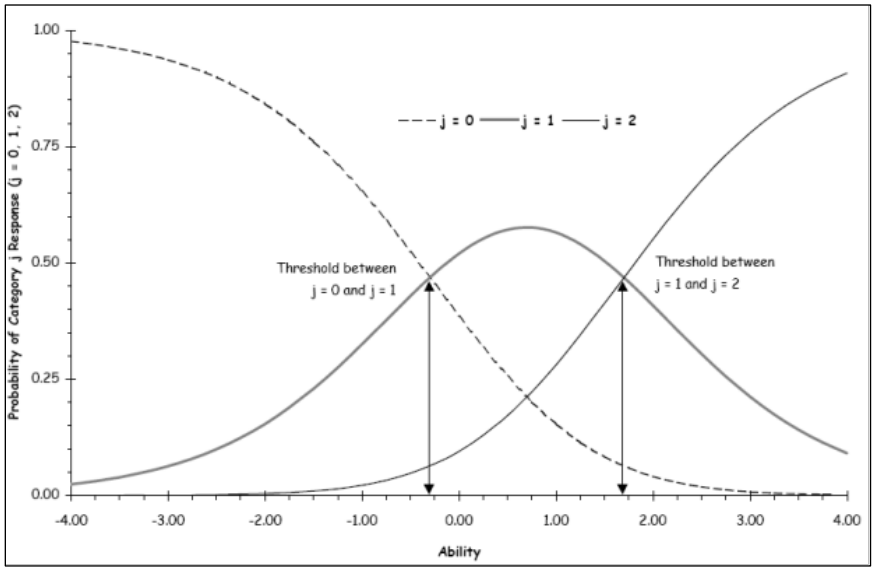


Figure 10.3 presents IRCs of obtaining each score category for a 2-point polytomously scored item. The dotted line (j=0) shows the probability of getting a score of 0. Students with very low ability (e.g., below -2) are likely to be in this category, which students receiving 1 point tend to fall in the middle range of abilities (the thick solid line curve, j=1). The thin solid line curve (j=2) represents the probability for those receiving scores of 2. Very high-ability students are more likely to be in this category, but some students of average and low ability can also get full credit for the item.

The points at which the lines cross have a similar interpretation for dichotomous items. For abilities to the left of (or less than) the point at which the j=0 line crosses the j=1 line, indicated by the left arrow, the probability is greatest for a 0 response. To the right of (or above) this point and up to the point at which the j=1 and j=2 lines cross (marked by the right arrow), the most likely response is a 1. For abilities to the right of this point, the most likely response is a 2. The probability of scoring a 1 response (j=1) declines in both directions as ability decreases to the low extreme and increases to the high extreme. These points may be thought of as the difficulties of crossing the *thresholds* between categories.

Figure 10.3. Example IRC for a 2-point Item



10.3. Data Preparation

Prior to any analyses, several steps were completed as preparation: (a) verify the data file containing student responses and apply the exclusion rules; (b) complete a traditional item analysis (TRIAN) and adjudication, where applicable, on all items; and (c) create incomplete data matrices (IDMs). A TRIAN of all SR items was conducted prior to calibration using classical statistics to identify potential test administration and score issues, as exemplified in the sample TRIAN report in Figure 10.4. Items with one or more of the following characteristics were flagged. A list of flagged items was communicated to the assessment specialists for review and confirmation that the correct key had been applied.

- *P*-value <0.15
- Item-total score correlation < 0.10
- Incorrect option selected by more high-performing students (top 33%) than the keyed response
- Distractor *p*-value ≥ 40%
- Distractor-total score correlation > 0
- One or more score points earned by less than 5% of students

Figure 10.4. Example TRIAN Report

Item	Form	Key	Corr.	*	PV<15	A%	*	B%	*	C%	*	D%	*	Omit%	Ncount
1	ALL	B	0.49			11		**46**		24		17		3	6578
2	ALL	D	0.46			17		12		9		**59**		2	6560
3	ALL	B	0.40			16		**50**		16		12		**6**	6572
4	ALL	D	0.47			5		9		21		**63**		2	6605
5	ALL	C	0.40			3		19		**51**		26		2	6643
6	ALL	C	0.46			12		5		**78**		4		2	6614
7	ALL	A	0.30			**33**		36		15		13		3	6643
8	ALL	C	0.43			21		35		**35**		6		3	6646

All TE items and ELA SR items were put through an adjudication process. For each item, the frequency distribution of responses scored correctly was created, along with the frequency distribution of responses scored as incorrect. Assessment specialists reviewed each response in the frequency reports and indicated whether the response should be scored as correct. The assessment specialists' indications were then cross-referenced with how the responses are scored to confirm that scoring is accurate. Figure 10.5 presents a sample adjudication spreadsheet.

Figure 10.5. Example Adjudication Spreadsheet

Item ID	Func.	Item Response	Scored Response	Freq. Count	% of Total Freq.	Date 1st Reviewer	1st Reviewer Initials	Issue? (Y/N)	Description of Issue	Date 2nd Reviewer	2nd Reviewer Initials	Issue? (Y/N)	Description of Issue
Item1		A_A1:B_B2	2	28339	59								
Item1		A_A1	1	35	0								
Item1		A_A1:A_A2	1	3782	8								
Item1		A_A1:C_C2	1	4803	10								
Item1		A_A1:D_D2	1	970	2								
Item1		A_A2	0	1	0								
Item1		B_B1	0	12	0								
Item1		B_B1:A_A2	0	464	1								
Item1		B_B2	0	4	0								
Item1		C_C1	0	10	0								
Item1		C_C1:A_A2	0	501	1								
Item1		C_C1:B_B2	0	841	2								
Item1		C_C1:C_C2	0	582	1								
Item1		C_C2	0	1	0								
Item1		D_D1	0	10	0								
Item1		D_D1:A_A2	0	652	1								

10.4. Checking Model Assumptions

It is important to evaluate how the IRT models applied for CMAS fit the data because reported scale scores are derived from theta estimated under the IRT models. Two major assumptions are investigated: unidimensionality and item fit. If the IRT models fit the data and the model assumptions are met, calibration of test items places both items and students on a scale that is independent of any sample of students up to a linear transformation. Equating is used to determine and apply a scale transformation that allows for meaningful comparisons of student performance across different forms or administrations of the test.

10.4.1. Unidimensionality

An assumption under the IRT models is unidimensionality, that there is exactly one latent variable (e.g., mathematics proficiency) that an instrument intends to measure. This is a more traditional and strict definition of the unidimensionality assumption. On the other hand, essential unidimensionality, in which there is one dominant latent variable with some minor latent variable(s), is a more practically applicable assumption (Stout, 1990). A factor analysis was performed on the item response data for the CMAS assessments to analyze the number of dimensions the assessments appear to be measuring. Given that unidimensional IRT models are used for calibration and scaling, it is important that there be evidence to support their use.

Appendix J presents the Spring 2024 scree plots. For most of the assessments, one factor explained most of the variance, which supports the use of a unidimensional IRT model, although the ELA/CSLA scree plots do suggest that Reading and Writing are distinct subscores. The loadings for Factor 2 for ELA were all much higher for the PCR trait items than any other items. This may indicate the influence of a writing construct that is separate from what is measured by the reading items.

10.4.2. Item Fit

Appendix K presents the item fit results. Item fit refers to how well the data fit the IRT calibration model, and it is evaluated using Yen's (1981) Q_1 statistic that allows for the evaluation of an item's IRT model fit to observed student performance. In the calculations of Q_1, the observed and expected (based on the model) frequencies were compared at 10 intervals, or deciles, along the scale. Yen's Q_1 fit statistic was computed for each item using the following formula:

$$Q_{1_i} = \sum_{j=1}^{10} \frac{N_{ij}\left(O_{ij} - E_{ij}\right)^2}{E_{ij}\left(1 - E_{ij}\right)}$$

where N_{ij} is the number of students in interval j for item I, and O_{ij} and E_{ij} are the observed and expected proportions of students in interval j for item i. The Q_1 statistic was then transformed so that items with different degrees of freedom can have comparable fit statistics:

$$Z_{Q_{1_i}} = \frac{Q_{1_i} - df}{\sqrt{2df}}$$

where *df* is the degree of freedom for the statistic (*df* = 10–the number of parameters estimated; *df* = 7 for SR items in a 3PL model). If $Z_{Q_{1_i}}$ is greater than Z_{crit}, the item is flagged for poor model fit:

$$Z_{crit} = \frac{N_i * 4}{1500}$$

where N_i is the sample size.

10.5. Calibration

Calibration refers to the estimation of item parameters that places items and students on a common scale. The GPC model was applied to the CMAS items to obtain the item parameter estimates using IRTPRO, with all operational item parameters estimated in a single calibration (i.e., concurrent calibration) for each assessment. For CSLA, the RPCM was applied to all CSLA items to obtain item parameter estimates using Winsteps, with all operational items within a grade also calibrated concurrently.

PCR items were calibrated at the (unweighted) trait score level rather than as aggregated scores. To account for potential local dependence between the two trait scores, the item response matrix was modified before operational calibrations. For each PCR item, one of the two trait scores for each student was randomly selected, and the non-selected trait score was then removed from the dataset and treated as missing for calibration. The resulting item response dataset, known as a "Moulder" matrix, contained roughly half as many observations for each PCR trait score as for the non-PCR items. However, the datasets still contained an adequate number of student responses to conduct the calibrations. Due to the small population of students taking the CSLA assessment, trait scores were not removed from the data when conducting calibrations for CSLA.

10.6. Equating

Equating is used to place new test forms onto the operational base scale. Equating of the operational test forms involves adjusting for differences in the difficulty of forms, both within and across assessment administrations, to ensure that students taking one form of a test are neither advantaged nor disadvantaged when compared to students taking a different form. Each time a new form is constructed, equating is used to allow scores on the new form to be comparable to scores on the previous form.

10.6.1. Mathematics, ELA, and Science

The Spring 2024 CMAS Mathematics, ELA, and Science assessments were calibrated and post-equated to the base scale following the procedures described below. All post-equating analyses were conducted using a representative sample of students that was evaluated based on the following demographics to ensure that the expected population demographic distributions were met: gender, ethnicity/race, economic disadvantage, language proficiency, students with disabilities, and district setting. A common items approach was used for equating the operational forms. Forms from adjacent administrations contain a set of items that are the same across the two administrations. This set of items represents the blueprint in terms of content and represents roughly 40% of a full form.

10.6.1.1. Consistency of Constructed-Response Scoring Check

Because the ELA assessments include a high percentage of CR items, the anchor sets include CR items to be more reflective of the construct being measured. For accurate equating, it is important that the items in the anchor sets be consistently scored across administrations. With SR items, scoring is the same each time the item is administered (e.g., 'A' is always scored as the correct answer) such that changes in item performance across administrations can be solely attributed to changes in student performance. With CR items, scoring is done by human raters, so it is important that scoring be monitored both within an administration and across administrations to maintain consistent scoring throughout. Such procedures were in place, including consistency in training and the use of validity papers throughout scoring.

As an additional check, the consistency of the CR scoring was examined prior to equating via the rescoring of a subset of the previous year's papers to remove any items that exhibited statistical drift in scoring characteristics so that the accuracy of the equating was not jeopardized. If a CR item appeared to lack consistency across the administrations, considerations were given to removing the item from the anchor set.

10.6.1.2. Stability Check

The item parameter stability check for the anchor items was conducted using classical item analyses, scatterplots of item parameter estimates, and ICC comparison. For the ICC comparison, old and new ICCs were compared using the *z*-score approach based on D^2 (Wells et al., 2014), as outlined below:

1. Obtain the theoretically weighted estimated posterior theta distribution using 31 quadrature points (-5 to 5).
2. Compute the slope and intercept constants using the Stocking and Lord (1993) method with all anchor items in the linking set.
3. Place the original anchor item parameter estimates onto the baseline scale by applying the constants obtained in Step 2.
4. For each anchor item, calculate D^2 between the ICCs based on old (x) and new (y) parameters at each point in this theta distribution:

$$D_i^2 = \sum^{k} \left[P_{ix}(\theta_k) - P_{iy}(\theta_k)\right]^2 * g(\theta_k)$$

   where i = item, x = old form, y = new form, k = theta quadrature point, and g = theoretically weighted posterior theta distribution.
5. Flag the items with a D^2 greater than 0.004.

10.6.1.3. Calibration and Anchor Set Evaluation

The initial calibration results were reviewed for problematic item parameter estimates, and fit plots were examined to detect items with poor model-data fit. Review of anchor item stability analyses resulted in dropping 1–4 items from the anchor set depending on the grade. The final ELA anchor sets represented between 34% and 45% of the unweighted total test points, the final mathematics anchor set represented between 25% and 39% of the total test points, and the final science anchor set represented between 36% and 90% of the total test points (grade 11 was a subset of the 2023 form, so there were a greater number of common items than usual). The online and paper versions were constructed to be parallel, and item parameter estimates were assumed to be the same. The information provided for the item statistics and IRT curves are based on the online estimates.

10.6.1.4. Final Anchor Sets

Items flagged from the stability check and consistency of CR scoring check were examined, and consideration was given to the impact of flagged item(s) on the content representativeness of the resulting anchor set. A flag alone was not the sole criteria for removing an item from the linking set; it was important to also make sure that the remaining anchor set continued to be representative of the overall content and structure of the test.

10.6.1.5. Equating Method

Using the item parameter estimates for the anchor set from the item bank and the current administration, the computer program STUIRT was used to obtain the transformation constants to place the current administration's items on the operational scale using the Stocking and Lord (1983) method. The scale transformation constants, Slope A and Intercept B, were applied to the item parameter estimates to place the new test items (new, *N*) on the operational scale (old, *O*) (Kolen & Brennan, 2004), as follows:

$$\alpha_{jO} = \alpha_{jN} / A$$

$$b_{jO} = A * b_{jN} + B$$

$$d_{jvO} = A * d_{jvN}$$

10.6.1.6. Paper Forms

Online and paper items were developed to be parallel to the online items. Operational paper items deemed identical to the operational online items were assumed to have the same item parameter estimates. Paper items were fixed to their online counterparts' item parameter estimates. This process produced item parameter estimates for all paper items.

10.6.2. CSLA

A common items approach was used to equate the CSLA operational forms. Forms from adjacent administrations contained a set of items that were the same across the two administrations (i.e., anchor items). Anchor items were operational items already equated to the base scale. The anchor items were placed in the same positions across all test forms within a grade and anchored the scale between the new test form and the base scale. This set of items represents the blueprint in terms of content and represents roughly 30% of a full form.

10.6.2.1. Stability Check

The stability check for the CSLA anchor items was conducted using classical item analysis, scatter plots of item difficulty, and displacement estimates from Winsteps. Items were flagged if the absolute value of the displacement estimate was greater than or equal to 0.30.

10.6.2.2. Final Anchor Sets

Items flagged from the stability check were examined, and consideration was given to the impact of flagged item(s) on the content representativeness of the resulting anchor set. A flag alone was not the sole criteria for removing an item from the linking set. It was important to also make sure that the remaining anchor set continues to be representative of the overall content and structure of the test. The final anchor sets for grades 3 and 4 represented 36% and 46%, respectively, of the unweighted total test points.

10.6.2.3. Equating Method

To obtain equated Rasch parameter estimates for the assessments, anchor item parameter estimates for each grade-level assessment were fixed to their previously equated item parameter estimates before calibrating the remaining non-anchor operational items on that assessment. This method placed the non-anchor operational items on the same scale as the anchor items.

10.7. Field Test Equating

The field test equating process is similar to operational equating, except the anchor items are the operational items. This process places the field test item parameter estimates onto the operational base scale. All field test items are calibrated concurrently, except for the ELA PCR items.

A minimum of 3,000 student responses for each field tested PCR item per trait is sampled for scoring and calibration. Due to possible dependency between the two trait scores for each PCR item, the field test items on each ELA assessment went through two calibrations. The first calibration included all field test items except the Writing Knowledge Language and Conventions (WKL) trait scores, and the second calibration included all field test items except the Writing Written Expression (WE) trait scores (with all operational items serving as anchor items in both cases).

The estimates from each calibration were then equated to the base scale separately following the same procedures as the operational equating. Finally, the two sets of equated field test parameters were combined by adding the equated field test WKL trait estimates to the equated estimates from the first calibration. This "double-calibration" method allowed for separate calibration of the field test trait scores while reducing the number of field test responses that needed to be scored per trait. Using a "Moulder" calibration method (as in the operational item calibration) would have meant using scoring resources to score traits that were never actually used for calibration or scoring.

10.8. Ability Estimates

For CMAS, student ability was estimated using IRT pattern scoring based on student responses and the operational item parameter estimates for all students who met the relevant attemptedness criterion. Student ability was estimated at the overall test level, as well as for Reading on the ELA assessment. Estimates were obtained via the maximum likelihood method (MLE) applied within the ISE software program. Pattern scores use the student's individual item response pattern (overall or Reading claim) to determine their ability estimate, which may lead to different ability estimates for the same raw score.

For CSLA, student abilities were estimated for each grade-level assessment by conducting an anchored calibration of the operational items' item parameter estimates after the item parameter estimates were obtained for the CSLA operational items. Student abilities were calculated for the overall test and for Reading. To obtain student ability estimates for the overall test, all the operational items were included in the anchored calibration. To obtain student ability estimates for Reading, only the operational items representing the specific claim were included in the anchored calibration. The calibrations included the weighting of the PCR WE trait score. Student ability estimates were obtained via the joint maximum likelihood method (JMLE) applied within Winsteps.

10.9. Overall and Subscale Scale Scores

For CMAS, student ability estimates for the overall test were transformed to scale scores ranging from 650 to 850 using the same scaling transformations as the prior year's administrations. The student ability estimates for the subscores for CMAS Science were transformed to scale scores ranging from 400 to 600. For ELA/CSLA, the student ability estimates for Reading were transformed to scale scores ranging from 110 to 190. The following linear transformation was used to convert examinee theta estimates into scale scores where A and B are unique scaling constants for each subject/grade:

$$SS = A * \theta + B$$

After the scale scores were calculated, the lowest obtainable scale score (LOSS) and highest obtainable scale score (HOSS) were applied. LOSS and HOSS were set to 650 and 850, respectively, for the overall test scale. For the Reading scale, LOSS and HOSS were set to 110 and 190. For the subscores of CMAS Science, LOSS and HOSS were set to 400 and 600.

10.10. Item-Level IRT Statistics

Appendix K presents the item parameter estimates for each grade. The item numbers are merely identifiers and do not reflect the sequence of items as they were presented to students. The "Item Type" uses the coding of SR for selected-response, XI for technology-enhanced, and CR for constructed-response items. The "Model" refers to the IRT model under which the item was estimated (2PL, 3PL, GPC, or RPCM). The "A" column shows the item parameter estimate for discrimination, "B" for difficulty, and "D1" through "D7" for GPC or RPCM category threshold estimates. Not all item parameters apply to each item. For example, there are no category threshold estimates for 2PL items.

The last column of the ELA, mathematics, and science tables reflects whether an item was flagged for misfit based on Q_1 for those calibrated assessments. Several items in each grade were flagged for misfit. Misfit plots for all items were reviewed, and misfit statistics were compared with data from the previous administration. Based on these reviews, no additional items were removed due to misfit flags. The last two columns for CSLA reflect the infit and outfit statistics generated from Winsteps. Fit values were reviewed, and no items were removed due to misfit.

10.11. IRT Curves

Appendix L presents the test characteristic curves (TCCs), test information curves (TICs), and CSEM curves for both the overall scale scores and the Reading scale scores. The 2024 CMAS TCCs matched those from 2018 (ELA and mathematics) and 2023 (science and CSLA) in terms of shape and position. The 2024 TCCs were reviewed across the distribution and at the cuts to ensure the match between years. Colorado's established maximum TCC difference of 0.05 was also maintained between the 2018 and 2024 forms. The TCCs are provided in terms of expected percent correct rather than expected raw score. Along with the curves, each cut score for a given grade is indicated with a red vertical line, as are the cut scores for Reading. On the overall scale score TCCs for science, mathematics, and ELA, the vertical line at a scale score of 750 corresponds to the cut for *Met Expectations* for each assessment.

10.12. Comparability of Online and Paper Forms

The scale score distributions for students taking the online and paper CMAS Mathematics, Science, and ELA assessments were examined using a matched samples approach to investigate the extent to which the online and paper forms produced comparable scores. Multiple variables were used for determining the matched groups to result in "equal" groups of online and paper students. The matching variables included sex, race/ethnicity, free and reduced lunch status, language proficiency, IEP, and district setting, plus the prior year's overall test score.

Because science is not assessed in consecutive grade levels, the prior year's score did not come from science. Rather, the grade 4 mathematics score was used for grade 5 science and the grade 7 mathematics score was used for the grade 8 science assessment. There were an insufficient (<1,000) number of students who took the grade 11 CMAS Science assessment on paper to complete a comparability and mode analysis.

Scale score distributions of CMAS scores between the matched samples were compared to estimate the mode effect. To quantify the differences between the two distributions, the effect size of the differences between the two distributions was calculated as Cohen's *d* (Cohen, 1977) using the mean scale score from each group and the pooled standard deviation:

$$d = \frac{M_{group\ 1} - M_{group\ 2}}{SD_{pooled}}$$

Suggested interpretations of Cohen's *d* are as follows:

- 0.2 = a small effect size
- 0.5 = a medium effect size
- 0.8 = a large effect size

A threshold for a possible mode effect was set to an effect size of 0.1 or greater and a matched sample size of at least 1,000 students. The effect size was calculated for the mathematics and ELA assessments in each grade, and for science in grades 5 and 8. The results were presented to CDE who made the final decision on whether to make an adjustment for mode differences for each assessment.

Table 10.2 presents the mode effect sizes from the Spring 2024 administration. Based on evaluation of the effect sizes, mode adjustments were made for ELA grades 6–8, mathematics grade 6, and science grade 5. For assessments where an adjustment was deemed necessary, scores from the paper form were adjusted using a linear transformation to match the mean and standard deviation of the online form. The conversion was applied to the overall scores. For ELA, the conversion was also applied to the Reading score. For science, the conversion was also applied to the subscales. For the paper-based mathematics assessments from the prior administration, mode adjustments from that prior administration were applied to those forms.

Table 10.2. Online vs. Paper Comparability Mode Effect Sizes

Assessment	N	Effect Size
Mathematics 3	3,309	0.03
Mathematics 4	2,964	0.01
Mathematics 5	2,547	0.09
Mathematics 6	2,206	-0.15
Mathematics 7	2,156	-0.10
Mathematics 8	1,951	-0.07
ELA 3	3,628	-0.05
ELA 4	3,148	-0.10
ELA 5	2,865	0.09
ELA 6	2,272	-0.18
ELA 7	2,055	-0.10
ELA 8	1,861	-0.14
Science 5	1,926	0.11
Science 8	1,606	-0.06
Science 11	N/A	N/A

Note. N/A = not applicable. Comparability analyses were not conducted for science grade 11 because the n-count was less than 1,000.

Chapter 11: Reliability

The *Standards for Educational and Psychological Testing* (AERA et al., 2014) refer to reliability as the "consistency of scores across replications of a testing procedure" (p. 33). A reliable test produces stable scores; very similar score distributions would result if the test were administered repeatedly under similar conditions to the same students without memory or fatigue affecting the scores. The level of reliability/precision of scores has implications for validity. In other words, scores must be consistent and precise enough to be useful for intended purposes. If scores are to be meaningful, tests should produce stable scores if the same group of students were to take the same test repeatedly without any fatigue or memory of the test. The range of certainty around the score should also be small enough to support educational decisions. Reliability for the CMAS assessments is evaluated with the following analyses:

- Internal consistency (coefficient alpha)
- Standard error of measurement (SEM)
- Conditional standard error of measurement (CSEM)
- Decision consistent and accuracy
- Inter-rater agreement

11.1. Internal Consistency (Coefficient Alpha)

Within the framework of classical test theory, an observed test score is defined as the sum of a student's true score and error ($X = T + E$, where X = the observed score, T = the true score, and E = error). A true score is considered the student's true standing on the measure, while the error score reflects a random error component. Thus, error is the discrepancy between a student's observed and true score. Internal consistency is typically measured via correlations among the items on an assessment and provides an indication of how much the items measure the same general construct. High reliability of test scores implies that the test items within a subclaim are measuring a single construct, which is a necessary condition for validity when the intention is to measure a single construct.

The reliability coefficient of a measure is the proportion of variance in observed scores accounted for by the variance in true scores. The coefficient can be interpreted as the degree to which scores remain consistent over parallel forms of an assessment (Ferguson & Takane, 1989; Crocker & Algina, 1986). In the internal consistency method used to estimate reliability for the CMAS assessments, a single form is administered to the same group of students to determine whether students respond consistently across the items within a test. A basic estimate of internal consistency reliability is Cronbach's coefficient alpha statistic (Cronbach, 1951). Coefficient alpha is equivalent to the average split-half correlation based on all possible divisions of a test into two halves. Coefficient alpha can be used on any combination of dichotomous and polytomous test items and is computed as follows:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_{j=1}^{n} S_j^2}{S_X^2}\right)$$

where n is the number of items, S_j^2 is the variance of students' scores on item j, and S_X^2 is the variance of the total-test scores.

Coefficient alpha ranges from 0.0 to 1.0, where higher values indicate a greater proportion of observed score variance. Two factors affect estimates of internal consistency: test length and homogeneity of items. The longer the test, the more observed score variance is likely to be true score variance. The more similar the items, the more likely students will respond consistently across items within the test.

Table 11.1 – Table 11.4 present the coefficient alpha results overall and by subclaim for each content area. Appendix G presents the coefficient alpha estimates by demographic subgroup. The internal consistency values for the overall test ranged from 0.87 to 0.92. Given the differences in length, it is expected that the coefficient alpha for the overall test will be higher than that of the subscales.

The overall test reliability does not correspond directly with the overall student scale scores, as those are based on IRT pattern scoring. However, the overall estimates do provide evidence of unidimensionality of the assessments. Furthermore, the subgroup reliabilities were consistent for the various demographic subgroups, except for those based on language proficiency. The reliability of the tests tended to be lower for students identified as non-English proficient or limited English proficient.

Table 11.1. Coefficient Alpha by Subclaim—Mathematics

Grade	Overall	Subclaim A	Subclaim B	Subclaim C	Subclaim D
3	0.92	0.87	0.78	0.70	0.71
4	0.92	0.87	0.65	0.71	0.65
5	0.92	0.88	0.70	0.72	0.70
6	0.91	0.84	0.67	0.74	0.69
7	0.91	0.85	0.53	0.77	0.59
8	0.91	0.85	0.51	0.69	0.62

Table 11.2. Coefficient Alpha by Subclaim—ELA

Grade	Overall	Reading: Literary Text	Reading: Informational Text	Reading: Vocabulary	Writing: Written Expression	Writing: Knowledge and Use of Language Conventions	Reading	Writing
3	0.91	0.80	0.78	0.69	0.59	0.75	0.90	0.66
4	0.88	0.72	0.72	0.61	0.67	0.75	0.86	0.73
5	0.89	0.76	0.69	0.67	0.71	0.77	0.87	0.77
6	0.90	0.77	0.77	0.60	0.74	0.79	0.89	0.78
7	0.91	0.69	0.80	0.68	0.89	0.84	0.88	0.83
8	0.90	0.74	0.79	0.67	0.82	0.84	0.89	0.82

Table 11.3. Coefficient Alpha by Subclaim—CSLA

Grade	Overall	Reading: Literary Text	Reading: Informational Text	Reading: Vocabulary	Writing: Written Expression	Writing: Knowledge and Use of Language Conventions	Reading	Writing
3	0.87	0.71	0.57	0.69	0.64	0.74	0.85	0.81
4	0.90	0.84	0.66	0.66	0.70	0.75	0.88	0.84

Table 11.4. Coefficient Alpha by Content Standard—Science

Grade	Overall	Physical Science	Life Science	Earth and Space Science
5	0.90	0.81	0.66	0.77
8	0.92	0.76	0.83	0.76
11	0.88	0.73	0.70	0.70

Note. For grade 5, the content standard is Physical Science/Life Science.

11.2. Standard Error of Measurement (SEM)

The SEM is another measure of reliability. This statistic uses the standard deviation of test scores along with a reliability coefficient (e.g., coefficient alpha) to estimate the number of score points that a student's test score would be expected to vary if the student was tested multiple times with equivalent forms of the assessment. It is calculated as follows:

$$SEM = s_x\sqrt{1 - p_{XX}}$$

where S_x is the standard deviation of test scores, and p_{xx} is the reliability coefficient. There is an inverse relationship between the reliability coefficient and SEM: the higher the reliability, the lower the SEM.

Table 11.5 – Table 11.8 present the SEM results by subclaim for each content area. The classical SEM estimate is not reported for the overall test scale scores and the Reading subscore, as those scores are based on IRT pattern scoring rather than the sum of item scores.

Table 11.5. SEM by Subclaim—Mathematics

Grade	Subclaim A	Subclaim B	Subclaim C	Subclaim D
3	1.89	1.16	1.47	1.31
4	2.11	1.12	1.64	1.41
5	2.12	1.23	1.39	1.46
6	1.96	1.30	1.43	1.34
7	2.03	1.29	1.51	1.52
8	2.12	1.39	1.45	1.22

Table 11.6. SEM by Subclaim—ELA

Grade	Reading: Literary Text	Reading: Informational Text	Reading: Vocabulary	Writing: Written Expression	Writing: Knowledge and Use of Language Conventions	Writing
3	1.89	1.78	1.60	1.71	0.78	2.09
4	2.15	2.17	1.40	2.38	0.70	2.75
5	2.13	2.09	1.40	2.26	0.68	2.61
6	2.12	2.18	1.51	2.53	0.74	3.02
7	2.97	2.35	1.66	1.35	1.06	2.32
8	2.18	2.28	1.71	2.37	2.37	4.14

Table 11.7. SEM by Subclaim—CSLA

Grade	Reading: Literary Text	Reading: Informational Text	Reading: Vocabulary	Writing: Written Expression	Writing: Knowledge and Use of Language Conventions	Writing
3	2.01	1.79	1.56	0.88	0.81	1.21
4	2.05	1.92	1.36	1.03	0.82	1.30

Table 11.8. SEM by Content Standard—Science

Grade	Physical Science	Life Science	Earth and Space Science
5	1.94	1.57	2.07
8	1.86	2.19	1.84
11	1.77	1.58	1.43

Note. For grade 5, the content standard is Physical Science/Life Science.

11.3. Conditional Standard Error of Measurement (CSEM)

While the SEM provides an estimate of precision for an assessment, conditional standard error of measurement (CSEM) gives an indication of how measurement error varies across the score scale. While coefficient alpha is reported as a measure of internal consistency of the items that each scale comprises, IRT-based CSEM is a more appropriate measure of the measurement error associated with these scale scores because the reported scale scores for both the overall test and Reading are determined using IRT pattern scoring.

The CSEM is defined as the standard deviation of observed scores given a particular true score and is estimated within the IRT framework as the inverse of the test information function. Plots of test information curves (TICs) and CSEM across the score scale range are provided in Appendix L for both the overall scale scores and Reading scores.

Each scale score has a CSEM estimate that indicates what the most likely range of scores would be for students receiving that score if they tested multiple times. The CMAS assessments measure more accurately at a scale score near the middle of the scale than at the ends of the scale. During test construction, CSEMs are reviewed to ensure that they are minimized around the performance level cut scores.

11.4. Decision Consistency and Accuracy

The CMAS Mathematics and ELA/CSLA scales are divided into five performance levels that a student is placed in based on their scale score: *Did Not Yet Meet Expectations*, *Partially Met Expectations*, *Approached Expectations*, *Met Expectations*, and *Exceeded Expectations*. The consistency of a decision refers to the extent to which the same classification would result if a student were to take two parallel forms of the same assessment. However, since test-retest data are not available, psychometric models can be used to estimate the decision consistency based on test scores from a single administration. The accuracy of a decision refers to the agreement between a student's observed score classification and a student's true score classification if a student's true score could be known.

Procedures developed by Livingston and Lewis (1995) were used to estimate the consistency and accuracy of performance level classifications. For the overall test, consistency and accuracy estimates, along with PChance (i.e., the probability of a consistent classification due to chance) and Cohen's Kappa (κ) coefficient (Cohen, 1960), are calculated as follows:

$$K = \frac{P - P_c}{1 - P_c}$$

where P is the probability of consistent classification, and P_c is the probability of consistent classification by chance (Lee et al., 2000).

Table 11.9 presents the kappa interpretations. Table 11.10 presents the decision consistency and accuracy results, and Table 11.11 and Table 11.12 present the consistency and accuracy estimates at each cut score.

Table 11.9. Kappa Values

Value of Kappa	Strength of Agreement
< 0.20	Poor
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Good
0.81 – 1.00	Very Good

Table 11.10. Decision Consistency and Accuracy Estimates

Assessment	Accuracy	Consistency	PChance	Kappa
Mathematics 3	0.74	0.64	0.23	0.53
Mathematics 4	0.75	0.66	0.24	0.56
Mathematics 5	0.75	0.65	0.24	0.55
Mathematics 6	0.74	0.64	0.24	0.53
Mathematics 7	0.75	0.65	0.25	0.54
Mathematics 8	0.72	0.62	0.23	0.51
ELA 3	0.73	0.64	0.26	0.52
ELA 4	0.68	0.58	0.23	0.45
ELA 5	0.73	0.64	0.28	0.50
ELA 6	0.73	0.63	0.25	0.51
ELA 7	0.71	0.61	0.21	0.50
ELA 8	0.70	0.60	0.23	0.48
CSLA 3	0.72	0.61	0.27	0.47
CSLA 4	0.74	0.64	0.26	0.52
Science 5	0.78	0.70	0.32	0.56
Science 8	0.81	0.74	0.34	0.61
Science 11	0.75	0.67	0.34	0.50

Table 11.11. Accuracy of Cut Scores

Assessment	*Partially Met Expectations* Cut	*Approached Expectations* Cut	*Met Expectations* Cut	*Exceeded Expectations* Cut
Mathematics 3	0.95	0.92	0.91	0.89
Mathematics 4	0.94	0.91	0.92	0.95
Mathematics 5	0.95	0.91	0.92	0.92
Mathematics 6	0.94	0.90	0.92	0.95
Mathematics 7	0.95	0.91	0.92	0.95
Mathematics 8	0.91	0.91	0.92	0.93
ELA 3	0.94	0.92	0.91	0.95
ELA 4	0.94	0.90	0.89	0.88
ELA 5	0.96	0.91	0.89	0.92
ELA 6	0.96	0.92	0.90	0.91
ELA 7	0.95	0.92	0.91	0.80
ELA 8	0.94	0.91	0.90	0.88
CSLA 3	0.93	0.88	0.91	0.99
CSLA 4	0.93	0.90	0.92	0.98
Science 5	–	0.92	0.89	0.96
Science 8	–	0.93	0.88	0.99
Science 11	–	0.91	0.85	0.99

Table 11.12. Consistency of Cut Scores

Assessment	*Partially Met Expectations* Cut	*Approached Expectations* Cut	*Met Expectations* Cut	*Exceeded Expectations* Cut
Mathematics 3	0.93	0.89	0.88	0.87
Mathematics 4	0.91	0.88	0.89	0.94
Mathematics 5	0.92	0.88	0.88	0.91
Mathematics 6	0.91	0.87	0.89	0.94
Mathematics 7	0.93	0.87	0.89	0.94
Mathematics 8	0.88	0.87	0.89	0.92
ELA 3	0.91	0.88	0.87	0.93
ELA 4	0.91	0.86	0.85	0.86
ELA 5	0.94	0.88	0.85	0.91
ELA 6	0.94	0.88	0.86	0.89
ELA 7	0.92	0.88	0.87	0.78
ELA 8	0.91	0.87	0.86	0.86
CSLA 3	0.90	0.84	0.88	0.98
CSLA 4	0.91	0.87	0.89	0.96
Science 5	–	0.89	0.85	0.95
Science 8	–	0.90	0.84	0.99
Science 11	–	0.88	0.79	0.99

11.5. Inter-Rater Agreement

For CR items, inter-rater agreement examines the extent to which students would obtain the same score if scored by different scorers. For each operational item, 10% of the responses were scored by a second reader, which allowed for rater agreement statistics to be calculated. 0 presents the inter-rater agreement statistics for the CR operational items (i.e., the percentage of operational items with exact agreement, adjacent agreement, and non-adjacent agreement). The target exact plus adjacent agreement rate is 95% for all items. The following agreement rates were calculated for each CR item and presented in Appendix M:

- Exact agreement, which represents exact agreement between the two raters
- Adjacent agreement, which represents adjacent agreement between the two raters (i.e., a difference of 1 score points)
- Non-adjacent agreement, which represents a difference of more than 1 score point between the two raters

For the PCR items, the following additional analyses were also conducted:

- Quadratic kappa, which is a comparison between the mean square error of rating pairs that are supposed to agree (X_1, Y_1) and those that are unrelated (X_1, Y_2):

$$KAPPA = \frac{E([X_1 - Y_1]^2)}{E([X_1 - Y_2]^2)}$$

- Standardized mean differences:

$$\bar{Z} = \frac{\left|\bar{X}_{R_1} - \bar{X}_{R_2}\right|}{\sqrt{\frac{sd_{R_1}^2 + sd_{R_2}^2}{2}}}$$

- Correlations:

$$r = \frac{\sum(X_{i_{R_1}} - \bar{X}_{R_1})\,(X_{i_{R_2}} - \bar{X}_{R_2})}{\sqrt{\sum(X_{i_{R_1}} - \bar{X}_{R_1})^2 \sum\left(X_{i_{R_2}} - \bar{X}_{R_2}\right)^2}}$$

Chapter 12: Validity

"Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (AERA et al., 2014). As such, it is not the CMAS assessments that are validated but rather the interpretations of the scores. The purpose of the CMAS assessments is to provide information about a student's level of mastery of the Colorado Academic Standards (CAS). Mastery of the standards in the elementary and middle school grades indicates that a student is on track to being college and career ready at each grade level. In support of these ends, this technical report has described processes that were implemented throughout the CMAS assessment cycle with validity and fairness considerations in mind. This chapter describes the various sources of validity evidence for CMAS as outlined in the *Standards for Educational and Psychological Testing* (AERA et al., 2014), often referencing other chapters and sections of this report.

12.1. Evidence Based on Test Content

Evidence based on the content of the assessment is supported by the degree of correspondence between test items and content standards. The degree to which the test measures what it claims to measure is known as construct validity. The CMAS assessments adhere to the principles of evidence-centered design, in which the standards to be measured (i.e., the CAS) are identified, and the performance a student needs to achieve to meet those standards is delineated in the evidence statements (ESs) or evidence outcomes (EOs). Test items are reviewed for adherence to universal design principles to maximize the participation of the widest possible range of students.

12.1.1. Test Development Process

The item development process is driven by targets at the ES or EO level. Before developing items, Pearson uses target spreadsheets to create an internal item development plan (IDP) aligned with the expectations of test design and with consideration of attrition rates at committee review and data review. The validity of a state assessment relies on the methodology that frames the development and design of the assessment. In support of that claim, Pearson upholds these considerations as the cornerstones of the CMAS item and test development:

- The item development process ensures that the mathematics and ELA/CSLA items align to the ESs and EOs and that the science items align to the EOs.
- IDPs were designed to produce and maintain a robust item bank; items are written to address the scope of measured standards, grade-level difficulties, and cognitive complexity.
- The item and test development processes promote the equivalency of the online and paper-based assessments.
- Items were developed with the intention of being administered on multiple testing platforms.
- Item and test development processes are compliant with industry standards.

Content is also aligned through the articulation of performance in the performance level descriptors (PLDs). At the policy level, the PLDs include policy claims about the educational achievement of students who attain a particular performance level, and a broad description of the grade-level knowledge, skills, and practices that students performing at a particular performance level are able to demonstrate. Those policy-level descriptors are the foundation for the subject- and grade-specific PLDs, which, along with the ES or EO framework, guide the development of the items and tasks.

Gathering construct validity evidence for the CMAS assessments is embedded in the process by which the test content is developed and validated. At each step in the test development process, educators, assessment experts, and bias and sensitivity experts were involved in review of text, items, and tasks for accuracy, appropriateness, and freedom from bias, as described in Chapter 3. In the early stages of development, Pearson conducted research studies to validate the item and task development approach. One such study focused on student task interaction and was designed to collect data on students' experience with the assessment tasks and technological functionalities, as well as the amount of time needed to answer each task. Pearson also conducted a rubric choice study that compared the functioning of two rubrics developed to score the ELA PCR tasks. Quantitative and qualitative evidence was collected to support the use of a condensed or expanded trait scoring rubric.

An important consideration when constructing test forms is recognition of items that may introduce construct-irrelevant variance. Such items should not be included on test forms to help ensure fairness to all student subgroups. Data reviews and content and bias reviews are held with Colorado educators to identify any issues with items before they are included on an operational test form. Accommodations were also made available based on individual need documented in the student's approved IEP or 504 Plan, as described in Section 5.5.

The CMAS operational test forms were carefully constructed to align with the test blueprints and specifications based on the CAS. Chapter 4 provides details on the construction of the operational assessment forms, which demonstrates that all test forms adhered to the same test design used in previous years or were previously used operationally.

12.1.2. 2023 Alignment Study

An independent alignment study was conducted by the Human Resources Research Organization (HumRRO) in 2023 to provide further evidence to support the claim that the content of the CMAS Science test items matches the intended content as specified in the 2020 CAS (Revivo et al., 2023). For the study, three panels (one per grade) of Colorado educators were convened to review the alignment between the CMAS Science items and the CAS. Every effort was made to recruit panels consisting of teachers reflecting the various demographic subgroups and regions across Colorado. HumRRO applied alignment criteria they developed that was approved by CDE. This procedure required the panelists to (a) provide cognitive complexity ratings for each item, (b) indicate the CAS best aligned to each item, and (c) indicate if each item aligned to an SEP or CCC.

Overall, the results of the study provide validity evidence to support the claim that the content of the CMAS Science test items matches the intended content as specified in the CAS and test blueprint. Across all grades, items represented the intended content and reflected the multidimensional nature of the CAS, although only grade 5 items reflected appropriate levels of cognitive complexity whereas grades 8 and 11 items narrowly missed the cognitive complexity criterion requirements. The results of the alignment study have been considered during the item development process for subsequent administrations.

12.2. Evidence Based on Internal Structure

Analyses of the internal structure of a test typically involve studies of the relationships among test items and/or test components (i.e., subclaims) in the interest of establishing the degree to which the items or components appear to reflect the construct on which a test score interpretation is based (AERA et al., 2014, p. 16). The term *construct* refers to the characteristics that a test is intended to measure; in the case of the CMAS assessments, the characteristics of interest are the knowledge and skills defined by the CAS.

The CMAS assessments provide a full summative test score, a Reading score, and scores for the science reporting categories. Percent of points earned is reported for Writing and mathematics and ELA subclaims. The goal of reporting at this level is to provide criterion-referenced data to assess the strengths and weaknesses of a student's achievement in specific components of each content area compared with other students taking the same assessment (for overall and percent of points earned scores) and students who took the assessment in prior years (for overall scores). This information can then be used for a variety of purposes as indicated in Section 1.3. Evidence based on internal structure is provided in the following sections of this technical report:

- Subclaim correlations (Section 9.4)
- Internal consistency (Section 11.1)
- Factor analysis (Section 10.4.1)

12.3. Evidence Based on Relationships to Other Variables

Correlations were calculated between the mathematics, ELA, and science assessments, as shown in Table 12.1. (The samples include only students with valid scores on both assessments.) These scores may be expected to have lower correlations if the tests are measuring distinct constructs. The correlations between the scale scores of the CMAS assessments ranged from 0.76 to 0.83; these values are also very close to the 2018 values.

Table 12.1. Correlations between CMAS Scale Scores

Comparison	Grade	N	Correlation
ELA & Math	3	54,377	0.79
	4	55,465	0.77
	5	55,648	0.76
	6	54,320	0.78
	7	53,044	0.78
	8	49,950	0.76
ELA & Science	5	55,189	0.83
	8	48,934	0.82
Math & Science	5	55,887	0.80
	8	49,435	0.77

12.4. Evidence Based on Response Processes

As noted in the *Standards for Educational and Psychological Testing* (AERA et al., 2014), additional support for a particular score interpretation or use can be provided by theoretical and empirical evidence indicating that students are using the intended response processes when responding to the items in a test. This type of evidence may be gathered from interacting with students to understand what processes underlie their item responses. Evidence may also be derived from feedback provided by test proctors/teachers involved in the administration of the test and raters involved in the scoring of CR items. Evidence may also be gathered by evaluating the correct and incorrect responses to short CR items (e.g., items requiring a few words to respond) or by evaluating the response patterns to multi-part items.

Prior to the 2016 administration, the PARCC consortium undertook research investigating the quality of the items, tasks, and stimuli, focusing on whether students interact with the online items/tasks as intended through cognitive labs. In these studies, students were asked to narrate how they interact with an item and answer questions about their experience with the item and online platform.

Cognitive labs were conducted for CMAS Science in 2024 to understand students' engagement with phenomenon-based assessment materials, both in terms of degree of engagement and the methods and cognitive processes by which students approach that engagement (Hendriksen, 2024). Pearson content experts conducted the cognitive interviews using a sample of one simulation and one static cluster set in grades 5 and 8. A total of 28 students participated. All interviews were conducted virtually and recorded to enable analysis of students' actions on the items (e.g., where a student clicks, how they move through the items), their non-verbal behavior, and transcription of the interviews. Students were asked to verbalize their thoughts out loud while reading and responding to the items. After each interview, interviewers completed a debriefing form to record additional thoughts on the interviews and note any issues if relevant. Most students expressed a preference for directions and could process the assessment stimuli and formulate their answer quickly and effectively, albeit not always correctly. Students' preferences of animations vs. static images were split, although their navigation behavior suggested that they more easily navigated the static clusters. The study resulted in recommendations from Pearson to improve the presentation and clarity of the directions and items. Appendix O presents the study report.

As new items are developed, the field test responses are reviewed. Sample responses to the CR items are also reviewed by educator committees during rangefinding to ensure that the rubrics make sense and provide example scored responses. During the data review meeting, item statistics are reviewed to ensure that the students are responding to items in the expected way. Low item item-total correlations and aberrant response distributions can all indicate that there are unexpected issues with either the correct or incorrect responses. Items where the correct response is not accurate or there are distractor responses that are technically correct can be identified and rejected at this step. During the adjudication step, incorrect responses to fill-in-the-blank items are also reviewed to make sure that no technically correct responses are excluded. These include entry issues such as extra spaces or unexpected responses such as adding an unnecessary decimal (e.g., 3.0 rather than 3).

12.5. Evidence Based on the Consequences of Testing

Because state tests are administered "in the expectation that some benefit will be realized from the intended use of the scores" (AERA et al., 2014, p. 19), validity evidence supporting the use and interpretation of CMAS results may be investigated as a consequence of testing. One intended consequence of testing is that more students will demonstrate mastery over the CAS over time, as evidenced by more students achieving in the top performance levels, if the data are used appropriately to make improvements in programming at the school and district levels. After the disruptions in education caused by the COVID-19 pandemic, the CMAS results have also served the purpose of monitoring student progress back to pre-pandemic levels.

Table 12.2 presents the percentage of students who have reached proficiency on the CMAS assessments over the years. Mathematics and ELA have been administered since Spring 2015, and CSLA has been administered since Spring 2016. Prior to 2018, students in grades 7 and 8 who were in advanced courses could take high school course-specific assessments. Starting in 2018, this was no longer allowed, which impacted the overall performance on the grades 7 and 8 mathematics assessments so the performance can only be compared starting with 2018. While science has been administered since Spring 2014, the change in standards and creation of a new assessment aligned to those standards means that performance cannot be compared before 2022.

As shown in the table, student performance has improved since the first administration except for grade 6 mathematics and grade 3 CSLA. Comparing 2024 performance to 2019, students in grades 3–5 have met or passed their 2019 performance in mathematics, and grades 3, 4, 6, and 7 have nearly met or passed their 2019 performance in ELA.

Table 12.2. Student Performance Over Time

Assessment	1st Admin. *%Met or Exceeded*	2019 *%Met or Exceeded*	2022 *%Met or Exceeded*	2023 *%Met or Exceeded*	2024 *%Met or Exceeded*	%Change, 1st Admin. to 2019	%Change, 1st Admin. to 2022	%Change, 1st Admin. to 2023	%Change, 1st Admin. to 2024
Mathematics 3	36.7	41.0	39.4	40.4	41.7	4.3	2.7	3.7	5.0
Mathematics 4	30.2	33.6	30.7	32.7	34.1	3.4	0.5	2.5	3.9
Mathematics 5	30.1	35.7	34.9	36.5	37.3	5.6	4.8	6.4	7.2
Mathematics 6	31.7	29.5	26.3	28.2	29.2	-2.2	-5.4	-3.5	-2.5
Mathematics 7	28.8	31.6	25.1	26.3	29.8	2.8	-3.7	-2.5	1.0
Mathematics 8	28.2	36.9	32.4	32.7	32.5	8.7	4.2	4.5	4.3
ELA 3	38.2	41.3	40.7	39.9	42.1	3.1	2.5	1.7	3.9
ELA 4	41.7	48.0	44.1	43.8	42.0	6.3	2.4	2.1	0.3
ELA 5	40.5	48.4	45.4	47.8	47.3	7.9	4.9	7.3	6.8
ELA 6	39.1	43.6	43.0	43.4	44.0	4.5	3.9	4.3	4.9
ELA 7	41.0	46.5	41.8	45.0	46.3	5.5	0.8	4.0	5.3
ELA 8	40.9	46.9	43.9	42.4	42.8	5.9	3.0	1.5	1.9
CSLA 3	22.0	27.5	19.8	18.7	17.2	5.5	-2.2	-3.3	-4.8
CSLA 4	13.9	19.1	13.7	14.2	15.9	5.2	-0.2	0.3	2.0
Science 5	33.9	–	–	33.9	37.9	–	–	–	4.0
Science 8	31.3	–	–	31.3	32.3	–	–	–	1.0
Science 11	24.6	–	–	24.6	24.8	–	–	–	0.2

Note. The first administration for mathematics grades 3–6 and ELA was Spring 2015, the first comparable administration for mathematics grades 7 and 8 was 2018, the first administration for which scale scores and performance levels were generated for science was Spring 2023, and the first administration for CSLA was Spring 2016. Performance results are not included for Spring 2021 due to low participation.

12.6. Fairness

Fairness is an important aspect of validity, as it is critical that an assessment provide accurate measurements for all students. To that end, the following fairness considerations were woven into the development and administration of the CMAS assessments:

- Sample items that provide the opportunity for teachers and students to become familiar with the test design and scoring of the assessments before experiencing the items on an operational test (Section 5.3)
- Universal design principles that are adhered to during the test development process with the goal of avoiding construct-irrelevant aspects of the assessment that could impact student performance (Chapter 3)
- Items are reviewed by educators for potential issues which could impact the performance of student groups prior to field testing (Chapter 3).
- Differential item functioning (DIF) analyses to identify any items that appear to be unfairly favoring one subgroup over another. All items which show DIF are reviewed by educators for potential bias in the item. (Section 3.7)
- Accessibility tools and accommodations to allow students to fully demonstrate their content knowledge without being hindered by non-construct related elements (Sections 4.3 and 5.5)

Participation information must also be reviewed and taken into consideration thoughtfully when interpreting the district and school results. As participation rates vary across student, school, and district groups, challenges with interpreting results increase. Depending on the specific school or district, some student groups may have been overrepresented in the results and others may have been underrepresented. Students may have also experienced ongoing reduced, disrupted, and/or adjusted learning opportunities during the school year. Due to these factors and many more challenges experienced due to COVID-19, districts and schools should be cautious when interpreting results because the data may not support all cross-state comparisons and historical uses when participation rates are low and/or representativeness is limited.

References

Achieve. (2019). *A framework to evaluate cognitive complexity in science assessments*. https://files.eric.ed.gov/fulltext/ED603610.pdf

Allen, N. L., Carlson, J. E., & Zelenak, C. A. (1999). *The NAEP 1996 technical report (NCES 1999–452).* National Center for Education Statistics, US Department of Education.

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

Beimers, J. N., Way, W. D., McClarty, K. L., & Miles, J. A. (2012, January). Evidence based standard setting: Establishing cut scores by integrating research evidence with expert content judgments. Pearson.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 392–479). Addison-Wesley.

Chien, M., & Shin, D. (2012). *IRT score estimation program*, V1.3 [computer program]. Pearson.

Cizek, G. J., Bunch, M. B., & Koons, H. (2004). Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice*, *23*(4), 31–31.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–47.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Academic Press.

Colorado Department of Education (CDE). (2016). *Colorado Spanish Language Arts (CSLA) standard setting report*. CDE and Pearson.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Harcourt Brace Jovanovich College Publishers.

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

Davis, L. L., & Moyer, E. L. (2015). *Performance level setting technical report*.

Dorans, N. J., & Holland, P. W. (1992). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning: Theory and practice* (pp. 35–66). Erlbaum.

Ferguson, G. A., & Takane, Y. (1989). *Statistical analysis in psychology and education* (6th ed.). McGraw-Hill.

Ferrara, S., Perie, M., & Johnson, E. (2008). Matching the judgmental task with standard setting panelist expertise: The item-descriptor (ID) matching method. *Journal of Applied Testing Technology, 9*(1), 1–22.

Hendriksen, M. (2024). *Analysis of the Colorado science assessment cognitive interviews*. Report prepared for the Colorado Department of Education by Pearson.

Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, *64*(3), 425–461.

Kim, S., & Kolen, M. (2004). *STUIRT [computer program].* The University of Iowa.

Kolen, M. J. & Brennan, R. L. (2004). *Test equating: Methods and practices.* (2nd ed.). Springer-Verlag.

Lee, W., Hanson, B. A., & Brennan, R. L. (2000, October). *Procedures for computing classification consistency and accuracy indices with multiple categories.* (ACT Research Report Series 2000–10). ACT, Inc.

Linacre, J. M. (2011). *Winsteps® Rasch measurement computer program* [computer program]. Winsteps.com.

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

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement,* 16, 159–176.

National Research Council (NRC). (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press. https://doi.org/10.17226/13165

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academic Press. https://www.nextgenscience.org/search-standards

Pearson. (2024). *Colorado Measures of Academic Success (CMAS) Science 2022 standard setting report*. Report developed for the Colorado Department of Education.

Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology, 19*, 49–57.

Revivo, R. Z., Dickinson, E. R., & Borawski, E. A. (2023, October). *Colorado Measures of Academic Success (CMAS) Science alignment study report*. Human Resources Research Organization (HumRRO).

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*(2), 201–210.

Stout, W. F. (1990). A new item response theory modelling approach and applications to unidimensionality assessment and ability estimation. *Psychometrika, 55*, 293–325.

Vector Psychometric Group. (2022). *IRTPRO* [computer program].

Wells, C. S., Hambleton, R. K., Kirkpatrick, R., & Meng, Y. (2014). An examination of two procedures for identifying consequential item parameter drift. *Applied Measurement in Education, 27,* 214–231.

Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice*, *31*(1), 2–13.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. MESA Press.

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. MESA Press.

Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement, 5*(2), 245–262.

Appendix A: Test Blueprints

The following tables present the percentage targets for each content area and grade-level assessment.

Table A.1. Test Blueprint—Mathematics Grades 3–5

Item Type/Subclaim/Calculator Use	Grade 3	Grade 4	Grade 5
Total #Points	50–51	50–51	50–51
Type I	61–62%	61–61%	61–62%
1.1	49–50%	37–38%	37–38%
1.2	12%	24%	24%
1.4	–	–	–
Subclaim A: Major Content	43–44%	47–48%	45–46%
Subclaim B: Supporting Content	18%	14%	16%
Type II	20–22%	20–22%	20–22%
2.3	6–12%	6–12%	6–12%
2.4	8–16%	8–16%	8–16%
Subclaim C: Expressing Mathematical Reasoning	20–22%	20–22%	20–22%
Type III	18%	18%	18%
3.3	6%	6%	6%
3.6	12%	12%	12%
Subclaim D: Modeling and Application	18%	18%	18%
Calculator	–	–	–
Non-Calculator	100%	100%	100%

Table A.2. Test Blueprint—Mathematics Grades 6–8

Item Type/Subclaim/Calculator Use	Grade 6	Grade 7	Grade 8
Total #Points	50–51	50–51	50–51
Type I	61–62%	61–62%	61–62%
1.1	37–38%	37–38%	33–34%
1.2	16%	16%	12–20%
1.4	8%	8%	8–16%
Subclaim A: Major Content	39–40%	45–46%	41–42%
Subclaim B: Supporting Content	22%	16%	20%
Type II	20–22%	20–22%	20–22%
2.3	6–12%	6–12%	6–12%
2.4	8–16%	8–16%	8–16%
Subclaim C: Expressing Mathematical Reasoning	20–22%	20–22%	20–22%
Type III	18%	18%	18%
3.3	6%	6%	6%
3.6	12%	12%	12%
Subclaim D: Modeling and Application	18%	18%	18%
Calculator	72–73%	76%	72–73%
Non-Calculator	27–28%	24%	27–28%

Table A.3. Test Blueprint—ELA Grades 3–5

Subclaim	Grade 3 (includes CSLA)	Grade 4 (includes CSLA)	Grade 5
Total #Points	53 (65)	59 (73)	57 (71)
Reading	77% (63%)	78% (63%)	77% (62%)
Literary Text	32% (26%)	31–34% (25–27%)	32% (25%)
Informational Text	26% (22%)	31% (25%)	32% (25%)
Vocabulary	19% (15%)	14–17% (11–14%)	14% (11%)
Writing	23% (37%)	22% (37%)	23% (38%)
Written Expression	11% (28%)	12% (29%)	12% (30%)
Knowledge and Use of Language Conventions	11% (9%)	10% (8%)	10% (8%)

Note. The numbers in parantheses bare based on weighted Written Expression scores. Scores may not add up as expected due to rounding.

Table A.4. Test Blueprint—ELA Grades 6–8

Subclaim	Grade 6	Grade 7	Grade 8
Total #Points	62 (78)	64 (80)	64 (80)
Reading	77% (62%)	78% (63%)	78% (63%)
Literary Text	26–29% (21–23%)	28% (23%)	28% (23%)
Informational Text	35% (28%)	34% (28%)	34% (28%)
Vocabulary	13–16% (10–13%)	16% (13%)	16% (13%)
Writing	23% (38%)	22% (38%)	22% (38%)
Written Expression	13% (31%)	13% (30%)	13% (30%)
Knowledge and Use of Language Conventions	10% (8%)	9% (8%)	9% (8%)

Note. The numbers in parantheses bare based on weighted Written Expression scores. Scores may not add up as expected due to rounding.

Table A.5. Test Blueprint—Science Grade 5

Standard/Item Type	%Total Test Score Points	#Points: Cluster	#Points: Mini Cluster	#Points: Standalone
Physical	35%	7–9	0–6	3–7
Physical/Life	24%	7–9	0–6	0–7
Earth and Space	41%	7–9	0–6	4–10
Science and Engineering Practices	65%–75%	–	–	–
SR and TE	53%	–	–	–
CR	47%	–	–	–
Total	100%	24–26	10–12	15

Table A.6. Test Blueprint—Science Grade 8 (Middle School)

Standard/Item Type	%Total Test Score Points	#Points: Cluster	#Points: Mini Cluster	#Points: Standalone
Physical	34%	7	6–12	0–6
Life	36%	7	6–12	1–7
Earth and Space	30%	7	6	3
Science and Engineering Practices	65%–74%	–	–	–
SR and TE	53%	–	–	–
CR	47%	–	–	–
Total	100%	21	24	16

Table A.7. Test Blueprint—Science Grade 11 (High School)

Standard/Item Type	%Total Test Score Points	#Points: Cluster	#Points: Mini Cluster	#Points: Standalone
Physical	46%	6	5–9	4–7
Life	32%	6	4–8	2–6
Earth and Space	31%	9	5–8	2–5
Science and Engineering Practices	65%–74%	–	–	–
SR and TE	53%	–	–	–
CR	47%	–	–	–
Total	100%	18	19	13

Appendix B: Science Cognitive Complexity Framework

Colorado Measures of Academic Success (CMAS) Science

Final Cognitive Complexity Framework, January 2023

COLORADO
Department of Education

Item Cognitive Complexity

Note: Examples provided are not intended to be comprehensive of all items meeting that descriptor.

Item Cognitive Complexity			
	Low - Single Dimension	**Medium - Two Dimensions**	**High - All Three Dimensions**
Item Dimensionality[1]	Item requires demonstration of only one dimension.	Item requires integration of two dimensions, described in the CAS Learning Progressions documents.	Item requires integration of three dimensions, described in the CAS Learning Progressions documents.
	Low - Heavy	**Medium - Moderate**	**High - Minimal**
Scaffolding/Support	The task prescribes a *fully specified* approach for responding. - All components are provided and commonplace; student does not need to infer or select from them: o formulas o Punnett squares o components of energy in a system o unbalanced chemical equations o labels to be applied to familiar models - Problem-solving steps are provided in the stimulus; student only executes them.	The task focuses student to *apply* an approach that is only partially specified. - Student is provided either a partial set of components for a routine task or an excess of components for a more complex task; student must either infer what is missing or discard what is irrelevant: o suggest original components to incorporate into an incomplete model o apply the correct formula(s) out of several available options o explain an observed instance using a specified concept o describe the missing steps from an incomplete investigational procedure - There is some uncertainty associated with the outcome of the scenario. - Some portion of the task is accomplished by way of provided problem-solving steps, but the student must choose or devise some portion of the process.	The task frames a situation that the student must *interpret to select or develop multiple steps* of an approach. - Student is presented with multiple informational inputs between which the relationship is not immediately obvious. - A high degree of uncertainty is associated with the outcome of the scenario. - Student selects or develops multiple problem-solving steps to complete a specific, structured, defined task or goal: o design an investigation o use listed materials to develop an original model o explain an observed instance using one or more unspecified concepts o determine and apply a sequence of formulas to solve a problem

Final CMAS Cognitive Complexity Framework 2023
Adapted from Achieve Cognitive Complexity in Science Assessments and Task Analysis Guide for Science

	Low - Minimal	Medium - Surface	High - Intensive
Sensemaking Sensemaking situation: students are provided material without obvious ties/connections to content (e.g., language of the standard); they must use their knowledge of the standard to explain what they see in the material.	- Task is answerable via rote knowledge connected to the phenomenon solely by context. - Item requires no engagement with the stimulus. - The student can correctly answer without addressing the central concept [mystery/puzzle] of the phenomenon. - Focused on identifying an answer, not on explaining phenomena. o identify the components of a familiar system without explaining their importance to the system. o identify the trend in a graph without using it to explain or predict anything.	- Making sense of a phenomenon or addressing a problem is necessary to accomplish at least a portion of the item, or answering the item is a strategic step toward a sensemaking goal. - Answer requires: o use of information, data, or a model to develop an explanation or argument o connection of multiple pieces of information. - Task asks for standards-based explanation of observations, but not the detailed relationships behind those observations: o determine which of several data sets correlates with the trend under observation o determine which portion of a system is most directly connected to the phenomenon o given data and a proposed cause for a phenomenon, provide support for that cause from the data	- Making sense of a phenomenon or addressing a problem is the fundamental source of challenge in the item. - Meaningful (valid, accurate, causal, etc.) information must be distinguished from other information through reasoning: o speculate a cause for an unusual observation and provide support for that cause from given data o determine corresponding trends between multiple data sets and evaluate for causation o notice patterns within data and connect them to the phenomenon under consideration o predict how a change to one part of a system will impact another part - Task requires use of pertinent standard knowledge to explain both observations and the detailed relationships behind those observations.

[1] - Disciplinary Core Ideas (DCI), Science and Engineering Practices (SEP), Crosscutting Concepts (CCC)

Overall Item Cognitive Complexity Rating

The overall cognitive complexity rating for the item follows a majority rules approach when comparing the ratings for dimensionality, scaffolding/support, and sensemaking.

- Item complexity is **High** if at least two of the three categories are rated at the highest level
- Item complexity is **Low** if two are rated at the lowest level.
- Item complexity is **Medium** in all other cases.

Final CMAS Cognitive Complexity Framework 2023
Adapted from Achieve Cognitive Complexity in Science Assessments and Task Analysis Guide for Science

Cluster Stimulus Cognitive Complexity

Note: Use the following **only for** the purpose of rating cluster and simulation stimuli.

Cluster Stimulus Complexity			
	1 - Low	**2 - Medium**	**3 - High**
Phenomenon/ Stimulus Material	- The task provides a problem or a phenomenon that students are already familiar with how to explain or solve. - Student is presented a simple, probably familiar situation/scenario and selects the appropriate, direct scientific explanation for the phenomenon. - Context is rudimentary or taken directly from the EO, Clarification Statement, or DCI. - Information is limited to that specifically needed to address the task.	- The scenario presents a relatively new phenomenon that students might have some familiarity with, which contains some specific uncertainty for tasks to focus on. - The scenario has multiple facets of information for students to interpret at a grade-appropriate level of sophistication. - Within the scenario there are explicit cues and/or scaffolding to focus students toward related tasks. - The provided components are sufficient for students to arrive at the appropriate scientific explanation for the phenomenon. - Context is substantial and goes beyond examples listed in the standards text.	- The scenario presents a new phenomenon or problem that o is at a level that "figuring out" would be real and authentic for students o is not immediately explainable by the student o likely involves multiple appropriate ways to engage and pursue the task - Connection of context to the standards is indirect or unobvious.

Final CMAS Cognitive Complexity Framework 2023
Adapted from Achieve Cognitive Complexity in Science Assessments and Task Analysis Guide for Science

Appendix C: Sample Student Performance Reports

Confidential Student Performance Report
Spring 2024

Colorado Measures of Academic Success

Student: **FIRSTNAME M. LASTNAME**
SASID: **9999999999** Birthdate: **MM/DD/YYYY**
School: **SAMPLE SCHOOL NAME (9999)**
District: **SAMPLE DISTRICT NAME (9999)**

Watch a video about this report!

English Language Arts — Grade 3

CMAS is the only assessment given to all Colorado students that measures what students should know and be able to do at the end of each grade. This report describes your student's understanding of Colorado's grade 3 English Language Arts expectations. Scan the QR code to see a video that will talk you through your student's report.

Your student's performance is shown as:

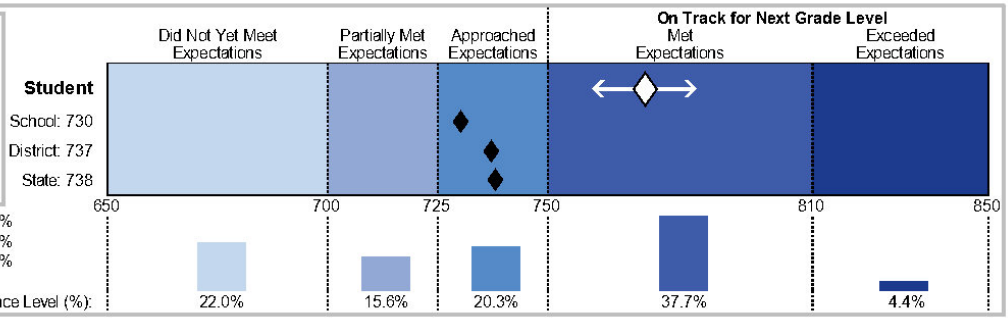
- A number on a scale between 650 and 850
- A performance level that is described below
- A percentile that shows how your student performed compared to other Colorado students

As you review this report:

- Review arrows around the large diamond to see where your student may have scored if the assessment was taken multiple times.
- Make school, district, and state comparisons with caution if participation is low.
- Talk with your student's teacher about your student's progress in English Language Arts.

Your Student's Score
772
Met Expectations
77th Percentile

School Participation: 98.4%
District Participation: 94.4%
State Participation: 94.2%



Performance Level Description - Met Expectations

3

FIRSTNAME **Met Expectations** and is on track for the next grade level. Students in this level typically demonstrate the following:

Reading
- Students understand easier 3rd grade texts in reading and may have a generally accurate understanding of more challenging texts.

Writing
- Students may effectively develop their ideas with evidence and organize their words almost always using correct spelling, punctuation, and capitalization, with few errors in grammar so that others can mostly understand their writing.

Knowledge and Use of Language and Conventions
- Students typically demonstrate command of the conventions of Standard English consistent with edited writing. Student writing includes errors in grammar and usage that may occasionally make understanding their writing difficult.

You can support your child at home by reading together and asking questions about what you read. Encourage your child to paraphrase what the story was about, tell what the story taught, and discuss how it relates to the child's experiences.

To view a video report and the full version of the performance level descriptor, visit **https://coassessments.com/parentsandguardians** or access the QR code.

Watch a video about this report!

Information about the Colorado Academic Standards measured by this assessment: **http://www.cde.state.co.us/coreadingwriting/statestandards**.

Page 1 of 2

04252024-Z99999999-5555-4444 - 0000000

FIRSTNAME M. LASTNAME

English Language Arts

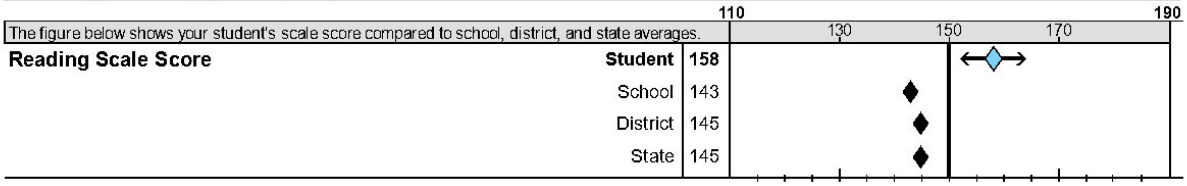
Confidential

Subclaim Performance

- The top diamond in the figure below shows your student's performance in Reading.
- The top bar in each of the other graphs shows the percent of points your student earned for writing and specific areas of reading and writing.
- District Averages are provided for comparison.
- State Averages are provided for comparison.
- Average of students at the Met Expectations performance level starting point.

Reading - Refer to page 1 for participation rates.

The figure below shows your student's scale score compared to school, district, and state averages.



Reading Scale Score	
Student	158
School	143
District	145
State	145

	Points Possible	Percent of Points Earned* (0% 25% 50% 75% 100%)
Literary Text Students read and analyze fiction, drama, and poetry.	17	71%
Informational Text Students read and analyze nonfiction, history, science, and the arts.	14	57%
Vocabulary Students use context to determine what words and phrases mean.	10	90%

	Points Possible	Percent of Points Earned* (0% 25% 50% 75% 100%)
Writing - Refer to page 1 for participation rates.		
Overall Writing Overall is calculated by multiplying Written Expression points by three and adding Language and Conventions points.	24	33%
Written Expression Students compose well-developed writing using details from what they have read.	6	33%
Language and Conventions Students demonstrate knowledge of conventions and other important elements of language.	6	33%

*Percent of points earned cannot be compared across years because individual test questions change from year to year. They also cannot be compared across specific areas of reading and writing because the number and difficulty of questions may not be the same.

For information about the CMAS assessment program, visit
http://www.cde.state.co.us/assessment/cmas.

Page 2 of 2

Confidential Student Performance Report
Spring 2024

Colorado Measures of Academic Success

Student: **FIRSTNAME M. LASTNAME**
SASID: **9999999999** Birthdate: **MM/DD/YYYY**
School: **SAMPLE SCHOOL NAME (9999)**
District: **SAMPLE DISTRICT NAME (9999)**

Watch a video about this report!

Mathematics — Grade 3

CMAS is the only assessment given to all Colorado students that measures what students should know and be able to do at the end of each grade. This report describes your student's understanding of Colorado's grade 3 Mathematics expectations. Scan the QR code to see a video that will talk you through your student's report.

Your student's performance is shown as:

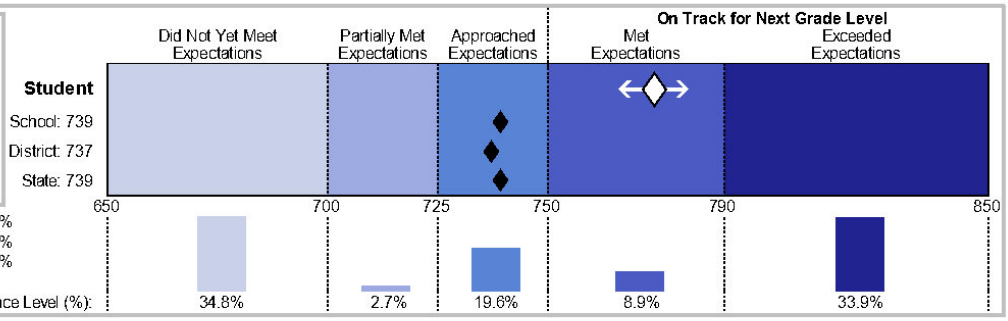
- A number on a scale between 650 and 850
- A performance level that is described below
- A percentile that shows how your student performed compared to other Colorado students

As you review this report:

- Review arrows around the large diamond to see where your student may have scored if the assessment was taken multiple times.
- Make school, district, and state comparisons with caution if participation is low.
- Talk with your student's teacher about your student's progress in Mathematics.

Your Student's Score
774
Met Expectations
58th Percentile

School Participation: 64.2%
District Participation: 64.8%
State Participation: 73.7%



Performance Level Description* - Met Expectations

3

FIRSTNAME034 **Met Expectations** and is on track for the next grade level. Students in this level typically demonstrate the following:

Major, Additional & Supporting Content:

- Find the missing numbers in problems where 1 factor is 5 or more.
- Show fractions with denominators 2, 4, and 8 on a number line, and use a picture to explain the relationship between fractions with the same denominator but different numerator, such as 2/4 and 3/4.
- Add and subtract to explain elapsed time. Measure and estimate liquid volume and mass. Show information on a picture graph, bar graph, or line plot with the correct units.
- Explain that the area inside a 2D shape is in square units. Solve problems to find unknown side lengths, and then find the perimeter of the shape. Explain the different types of four-sided shapes, such as squares, trapezoids, and rectangles, and what makes them different.

Expressing Mathematical Reasoning:

- Explain the correct way to solve a problem, without mistakes in calculation. Explain why the answer to a problem is correct or incorrect.

Modeling and Application:

- Estimate amounts in a real-world situation. Use the relationships between numbers to explain an answer. Make a model of a math problem, such as an expression.

To further support your student, you can work with your student on the following skills:

- Using mental math strategies to explain the relationship between multiplication and division in fact families
- Plotting and explaining values on a number line
- Providing an incorrect explanation of a math problem and asking your student to correct you and explain the student's thinking

Performance level descriptors (PLDs) are organized in a manner that assumes students demonstrating higher levels of command have mastered the concepts and skills within lower levels. To view a video report and the full version of the performance level descriptor, visit **https://coassessments.com/parentsandguardians** or access the QR code.

*Adapted from ilClassroom in Action's Performance Level Summaries

Watch a video about this report!

Information about the Colorado Academic Standards measured by this assessment:
http://www.cde.state.co.us/comath/statestandards.

04252024-Z9999999-9999-9999 - 0000000

FIRSTNAME M. LASTNAME

Mathematics

Confidential

Subclaim Performance

The top bar in each of the other graphs shows the percent of points your student earned for each of the four mathematics assessment subclaims.

District Averages are provided for comparison.

State Averages are provided for comparison.

Average of students at the Met Expectations performance level starting point.

	Points Possible	Percent of Points Earned* 0% 25% 50% 75% 100%
Mathematics - Refer to page 1 for participation rates.		
Major Content Students solve problems involving multiplication and division, area, measurement, and basic fraction understanding.	22	73%
Additional & Supporting Content Students solve problems involving perimeter, place value, geometric shapes, and representations of data.	9	67%
Expressing Mathematical Reasoning Students create and justify logical mathematical solutions and analyze and correct the reasoning of others.	11	64%
Modeling & Application Students solve real-world problems, represent and solve problems with symbols, reason quantitatively, and strategically use appropriate tools.	9	67%

*Percent of points earned cannot be compared across years because individual test questions change from year to year. They also cannot be compared across specific areas of math because the number and difficulty of questions may not be the same.

For information about the CMAS assessment program, visit **http://www.cde.state.co.us/assessment/cmas**.

Page 2 of 2

Confidential Student Performance Report
Spring 2024

Colorado Measures of Academic Success

Student: **FIRSTNAME M. LASTNAME**
SASID: **9999999999** Birthdate: **MM/DD/YYYY**
School: **SAMPLE SCHOOL NAME (9999)**
District: **SAMPLE DISTRICT NAME (9999)**

Watch a video about this report!

Science — Grade 5

CMAS is the only assessment given to all Colorado students that measures what students should know and be able to do at the end of each grade. This report describes your student's understanding of Colorado's grade 5 science expectations. Scan the QR code to see a video that will talk you through your student's report.

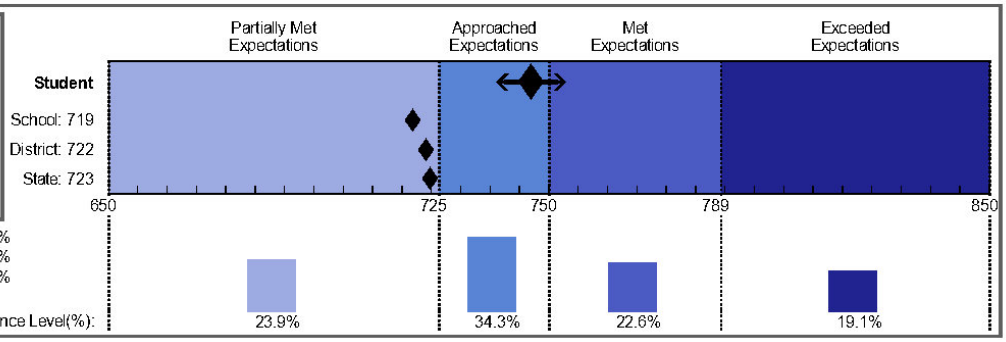
Your student's performance is shown as:

- A number on a scale between 650 and 850
- A performance level that is described below
- A percentile that shows how your student performed compared to other Colorado students

As you review this report:

- Review arrows around the large diamond to see where your student may have scored if the assessment was taken multiple times.
- Make school, district, and state comparisons with caution if participation is low.
- Talk with your student's teacher about your student's progress in science.

Your Student's Score
746
Approached Expectations
58th Percentile



School Participation: 94.4%
District Participation: 94.6%
State Participation: 89.8%
CO Students by Performance Level(%):

Performance Level Description - Approached Expectations

FIRSTNAME002 showed a moderate understanding of the Colorado Academic Standards' grade 5 science expectations and will likely need additional academic support in the next grade level. Students in the Approached Expectations level typically:

- Describe matter (particles too small to be seen) as always conserved, and mixing can result in new substances.
- Observe the properties of an object to identify it.
- Describe evidence that demonstrates Earth's gravity as the cause of objects being pulled toward its center.
- Show the transfer of energy from the Sun to things animals use as food.
- Describe matter and energy cycles in an ecosystem and explain that plants get materials to grow from air and water.
- Relate the distance between a star and Earth to the star's apparent brightness.
- Demonstrate Earth's patterns using shadows, day and night, and the seasonal appearance of some stars.
- Describe Earth's major systems and how they interact.
- Identify the proportions of salt water and fresh water in different reservoirs on Earth.
- Summarize ways that communities protect Earth's environment and resources.

5

To view a video report and the full version of the performance level descriptor, visit **https://coassessments.com/parentsandguardians/** or access the QR code.

Watch a video about this report!

Information about the Colorado Academic Standards measured by this assessment:
http://www.cde.state.co.us/coscience/statestandards.

Page 1 of 2 03192024-Z9999999-9999-9999 - 0000000

FIRSTNAME M. LASTNAME

Science

Confidential

Subscale Performance

- The shaded areas below represent about 70% of student scores across the state.
- Diamonds outside of the shaded area indicate a potential weakness or strength compared to the state.

Reporting Category Description	Subscale Score		Lower than Average / Average / Higher than Average (400 – 441 – 519 – 550)
Physical Science			
Common properties, forms, and changes in matter and energy	**494**	**Student**	
	460	School	
	463	District	
	464	State	
Physical/Life Science			440 – 522
Characteristics of living things, processes of life, and how living things interact with each other and their environment	**493**	**Student**	
	457	School	
	461	District	
	461	State	
Earth and Space Science			446 – 519
Processes and interactions of Earth's systems, and the structure and dynamics of Earth and other objects in space	**499**	**Student**	
	458	School	
	461	District	
	463	State	
Science and Engineering Practices			448 – 517
Making sense of the natural world through investigation and problem solving	**513**	**Student**	
	459	School	
	463	District	
	464	State	

Performance by Prepared Graduate Statements (PGs) and Grade Level Expectations (GLEs)

- PGs and GLEs identify what students need to master to be ready for the next grade level.
- The figure below shows the percent of points your student earned for each grade 5 science GLE.

Student's performance
District average
State average

Standard, PG and GLE	Points Possible	Percent of Points Earned* (0% 25% 50% 75% 100%)
Physical Science		
PG 1: Structure, properties, and interactions of matter		
GLE 1: Matter exists as particles too small to be seen; Properties can be used to identify materials	6	67%
GLE 2: Chemical reactions and the Law of Conservation of Mass	6	17%
GLE 3: Gravity	6	50%
Physical/Life Science		
PG 1: Structure, properties, and interactions of matter		
GLE 4: Energy from food was once energy from the sun		
PG 6: How living systems interact with the environment	6	50%
GLE 2: Plants get most of their material for growth from air and water		
PG 6: How living systems interact with the environment		
GLE 1: Matter cycles between air and soil; Organisms live and die	6	33%
Earth and Space Science		
PG 9: The universe and Earth's place in it		
GLE 1: Earth's major systems interact in multiple ways	8	38%
GLE 2: Interactions between Earth's orbit and the moon's orbit		
PG 10: How and why Earth is constantly changing		
GLE 3: Earth's major systems interact in multiple ways	7	71%
GLE 4: Earth's major water is in the ocean and much of Earth's freshwater is in glaciers or underground		
GLE 5: Societal activities have major effects on land, ocean, atmosphere and even outer space	6	50%

*Percent of points earned cannot be compared across years because individual test questions change from year to year. They also cannot be compared across PGs because the number and difficulty of questions may not be the same.

Page 2 of 2

Appendix D: Student Participation by Demographic Group

Table D.1. Student Participation by Demographics—Mathematics

Subgroup	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
No IEP	49,508	49,876	49,240	48,555	47,852	45,360
IEP	7,221	7,519	7,361	6,569	6,050	5,443
No Accommodation	52,386	52,526	51,567	50,434	49,322	46,544
Accommodation	4,343	4,869	5,034	4,690	4,580	4,259
Am. Indian/Alaska Native	344	350	378	338	333	331
Asian	1,958	2,025	2,108	1,954	1,903	1,665
Black	2,605	2,636	2,559	2,400	2,340	2,221
Hispanic	19,829	19,691	19,900	19,561	19,498	19,153
White	28,527	29,061	28,214	27,683	26,699	24,672
Hawaiian/Pacific Islander	187	210	214	196	204	171
Two or More Races	3,104	3,239	3,050	2,856	2,780	2,469
Missing	175	183	178	136	145	121
No Economic Disadvantage	31,978	32,536	32,240	31,658	31,222	29,028
Economic Disadvantage	24,751	24,859	24,361	23,466	22,680	21,775
Female	28,056	28,216	27,712	26,651	25,993	24,277
Male	28,663	29,164	28,878	28,456	27,892	26,512
Nonbinary	*	*	*	17	17	*
Language Proficiency NA	46,010	46,679	45,854	44,616	43,776	41,801
Language Proficiency NEP	3,370	2,767	1,916	1,679	1,823	1,851
Language Proficiency LEP	6,328	5,679	5,204	4,124	3,969	3,784
Language Proficiency FEP	1,021	2,270	3,627	4,705	4,334	3,367
Not Migrant	56,518	57,196	56,391	54,932	53,736	50,625
Migrant	211	199	210	192	166	178

*n-count less than 16

Table D.2. Student Participation by Demographics—ELA

Subgroup	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
No IEP	47,599	48,361	48,625	48,007	47,266	44,883
IEP	6,978	7,349	7,328	6,555	6,068	5,458
No Accommodation	50,853	51,350	51,186	50,130	48,985	46,328
Accommodation	3,724	4,360	4,767	4,432	4,349	4,013
Am. Indian/Alaska Native	344	348	380	334	332	331
Asian	1,931	1,989	2,070	1,934	1,880	1,641
Black	2,575	2,619	2,536	2,380	2,335	2,221
Hispanic	17,888	18,144	19,344	19,077	18,970	18,741
White	28,376	28,989	28,169	27,638	26,690	24,648
Hawaiian/Pacific Islander	189	208	215	193	202	169
Two or More Races	3,101	3,233	3,059	2,868	2,781	2,461
Missing	173	180	180	138	144	129
No Economic Disadvantage	31,345	32,059	31,950	31,408	30,973	28,808
Economic Disadvantage	23,232	23,651	24,003	23,154	22,361	21,533
Female	27,032	27,410	27,411	26,393	25,739	24,083
Male	27,535	28,285	28,530	28,151	27,578	26,244
Nonbinary	*	*	*	18	17	*
Language Proficiency NA	45,930	46,664	45,890	44,640	43,829	41,849
Language Proficiency NEP	2,046	1,645	1,237	1,104	1,238	1,340
Language Proficiency LEP	5,580	5,130	5,192	4,106	3,944	3,783
Language Proficiency FEP	1,021	2,271	3,634	4,712	4,323	3,369
Not Migrant	54,395	55,537	55,763	54,382	53,185	50,181
Migrant	182	173	190	180	149	160

*n-count less than 16

Table D.3. Student Participation by Demographics—CSLA

Subgroup	Grade 3	Grade 4
No IEP	1,391	1,029
IEP	169	133
No Accommodation	1,345	989
Accommodation	215	173
Am. Indian/Alaska Native	–	–
Asian	–	–
Black	–	*
Hispanic	1,550	1,154
White	*	*
Hawaiian/Pacific Islander	–	–
Two or More Races	–	–
Missing	–	–
No Economic Disadvantage	336	241
Economic Disadvantage	1,224	921
Female	762	582
Male	797	580
Nonbinary	*	–
Language Proficiency NA	–	–
Language Proficiency NEP	817	626
Language Proficiency LEP	743	536
Language Proficiency FEP	–	–
Not Migrant	1,543	1,152
Migrant	17	*

*n-count less than 16

Table D.4. Student Participation by Demographics—Science

Subgroup	Grade 5	Grade 8	Grade 11
No IEP	48,828	44,581	31,019
IEP	7,257	5,333	3,089
No Accommodation	51,523	46,136	31,394
Accommodation	4,562	3,778	2,714
Am. Indian/Alaska Native	374	324	235
Asian	2,094	1,653	970
Black	2,529	2,177	1,618
Hispanic	19,709	18,818	14,540
White	27,973	24,217	15,087
Hawaiian/Pacific Islander	210	168	115
Two or More Races	3,017	2,431	1,459
Missing	179	126	84
No Economic Disadvantage	31,999	28,543	19,199
Economic Disadvantage	24,086	21,371	14,909
Female	27,441	23,824	16,007
Male	28,635	26,076	18,064
Nonbinary	*	*	37
Language Proficiency NA	45,449	41,054	29,284
Language Proficiency NEP	1,869	1,803	1,023
Language Proficiency LEP	5,158	3,735	2,259
Language Proficiency FEP	3,609	3,322	1,542
Not Migrant	55,876	49,739	33,968
Migrant	209	175	140

*n-count less than 16

Appendix E: Scale Score Distributions

Table E.1. Scale Score Distribution—Mathematics Grade 3

SS	Freq.	%	Cum. Freq.	Cum. %
650	348	0.61	348	0.61
651	19	0.03	367	0.65
652	50	0.09	417	0.74
653	17	0.03	434	0.77
654	50	0.09	484	0.85
655	24	0.04	508	0.90
656	38	0.07	546	0.96
657	53	0.09	599	1.06
658	43	0.08	642	1.13
659	46	0.08	688	1.21
660	61	0.11	749	1.32
661	69	0.12	818	1.44
662	81	0.14	899	1.58
663	58	0.10	957	1.69
664	86	0.15	1,043	1.84
665	74	0.13	1,117	1.97
666	91	0.16	1,208	2.13
667	80	0.14	1,288	2.27
668	106	0.19	1,394	2.46
669	118	0.21	1,512	2.67
670	122	0.22	1,634	2.88
671	132	0.23	1,766	3.11
672	125	0.22	1,891	3.33
673	131	0.23	2,022	3.56
674	133	0.23	2,155	3.80
675	136	0.24	2,291	4.04
676	153	0.27	2,444	4.31
677	197	0.35	2,641	4.66
678	162	0.29	2,803	4.94
679	182	0.32	2,985	5.26
680	188	0.33	3,173	5.59
681	238	0.42	3,411	6.01
682	239	0.42	3,650	6.43
683	199	0.35	3,849	6.78
684	232	0.41	4,081	7.19
685	232	0.41	4,313	7.60
686	270	0.48	4,583	8.08
687	234	0.41	4,817	8.49
688	233	0.41	5,050	8.90
689	299	0.53	5,349	9.43
690	294	0.52	5,643	9.95
691	282	0.50	5,925	10.44
692	300	0.53	6,225	10.97
693	315	0.56	6,540	11.53
694	314	0.55	6,854	12.08
695	312	0.55	7,166	12.63
696	343	0.60	7,509	13.24
697	339	0.60	7,848	13.83
698	363	0.64	8,211	14.47
699	393	0.69	8,604	15.17
700	391	0.69	8,995	15.86
701	374	0.66	9,369	16.52
702	396	0.70	9,765	17.21
703	388	0.68	10,153	17.90
704	370	0.65	10,523	18.55
705	403	0.71	10,926	19.26
706	388	0.68	11,314	19.94
707	429	0.76	11,743	20.70
708	428	0.75	12,171	21.45
709	424	0.75	12,595	22.20
710	418	0.74	13,013	22.94
711	463	0.82	13,476	23.76
712	457	0.81	13,933	24.56
713	499	0.88	14,432	25.44
714	472	0.83	14,904	26.27
715	436	0.77	15,340	27.04
716	465	0.82	15,805	27.86
717	438	0.77	16,243	28.63
718	442	0.78	16,685	29.41
719	483	0.85	17,168	30.26
720	464	0.82	17,632	31.08
721	447	0.79	18,079	31.87
722	486	0.86	18,565	32.73
723	479	0.84	19,044	33.57
724	515	0.91	19,559	34.48
725	494	0.87	20,053	35.35
726	517	0.91	20,570	36.26
727	498	0.88	21,068	37.14
728	533	0.94	21,601	38.08
729	529	0.93	22,130	39.01
730	536	0.94	22,666	39.95
731	525	0.93	23,191	40.88
732	520	0.92	23,711	41.80
733	488	0.86	24,199	42.66
734	536	0.94	24,735	43.60
735	537	0.95	25,272	44.55
736	562	0.99	25,834	45.54
737	531	0.94	26,365	46.48
738	538	0.95	26,903	47.42
739	577	1.02	27,480	48.44
740	535	0.94	28,015	49.38
741	551	0.97	28,566	50.36
742	570	1.00	29,136	51.36
743	575	1.01	29,711	52.37
744	580	1.02	30,291	53.40
745	537	0.95	30,828	54.34

SS	Freq.	%	Cum. Freq.	Cum. %
746	527	0.93	31,355	55.27
747	580	1.02	31,935	56.29
748	532	0.94	32,467	57.23
749	594	1.05	33,061	58.28
750	537	0.95	33,598	59.23
751	539	0.95	34,137	60.18
752	533	0.94	34,670	61.12
753	529	0.93	35,199	62.05
754	529	0.93	35,728	62.98
755	583	1.03	36,311	64.01
756	571	1.01	36,882	65.01
757	530	0.93	37,412	65.95
758	509	0.90	37,921	66.85
759	522	0.92	38,443	67.77
760	562	0.99	39,005	68.76
761	551	0.97	39,556	69.73
762	529	0.93	40,085	70.66
763	495	0.87	40,580	71.53
764	515	0.91	41,095	72.44
765	533	0.94	41,628	73.38
766	499	0.88	42,127	74.26
767	505	0.89	42,632	75.15
768	454	0.80	43,086	75.95
769	542	0.96	43,628	76.91
770	462	0.81	44,090	77.72
771	462	0.81	44,552	78.53
772	427	0.75	44,979	79.29
773	475	0.84	45,454	80.12
774	442	0.78	45,896	80.90
775	444	0.78	46,340	81.69
776	430	0.76	46,770	82.44
777	391	0.69	47,161	83.13
778	381	0.67	47,542	83.81
779	372	0.66	47,914	84.46
780	364	0.64	48,278	85.10
781	371	0.65	48,649	85.76
782	381	0.67	49,030	86.43
783	354	0.62	49,384	87.05
784	346	0.61	49,730	87.66
785	328	0.58	50,058	88.24
786	316	0.56	50,374	88.80
787	318	0.56	50,692	89.36
788	312	0.55	51,004	89.91
789	302	0.53	51,306	90.44
790	263	0.46	51,569	90.90
791	259	0.46	51,828	91.36
792	303	0.53	52,131	91.89
793	234	0.41	52,365	92.31
794	227	0.40	52,592	92.71
795	215	0.38	52,807	93.09
796	246	0.43	53,053	93.52
797	199	0.35	53,252	93.87
798	211	0.37	53,463	94.24
799	183	0.32	53,646	94.57
800	175	0.31	53,821	94.87
801	161	0.28	53,982	95.16
802	178	0.31	54,160	95.47
803	159	0.28	54,319	95.75
804	116	0.20	54,435	95.96
805	159	0.28	54,594	96.24
806	125	0.22	54,719	96.46
807	141	0.25	54,860	96.71
808	116	0.20	54,976	96.91
809	109	0.19	55,085	97.10
810	109	0.19	55,194	97.29
811	107	0.19	55,301	97.48
812	100	0.18	55,401	97.66
813	79	0.14	55,480	97.80
814	66	0.12	55,546	97.91
815	76	0.13	55,622	98.05
816	93	0.16	55,715	98.21
817	58	0.10	55,773	98.31
818	85	0.15	55,858	98.46
819	57	0.10	55,915	98.57
820	56	0.10	55,971	98.66
821	39	0.07	56,010	98.73
822	50	0.09	56,060	98.82
823	78	0.14	56,138	98.96
824	50	0.09	56,188	99.05
825	45	0.08	56,233	99.13
826	24	0.04	56,257	99.17
827	28	0.05	56,285	99.22
828	22	0.04	56,307	99.26
829	39	0.07	56,346	99.32
830	40	0.07	56,386	99.40
831	23	0.04	56,409	99.44
832	26	0.05	56,435	99.48
833	30	0.05	56,465	99.53
834	9	0.02	56,474	99.55
835	11	0.02	56,485	99.57
836	15	0.03	56,500	99.60
837	11	0.02	56,511	99.62
838	13	0.02	56,524	99.64
839	16	0.03	56,540	99.67
840	6	0.01	56,546	99.68
841	25	0.04	56,571	99.72
842	28	0.05	56,599	99.77
843	4	0.01	56,603	99.78
844	4	0.01	56,607	99.78
845	9	0.02	56,616	99.80
846	2	0.00	56,618	99.80
847	5	0.01	56,623	99.81
848	4	0.01	56,627	99.82
849	2	0.00	56,629	99.82
850	100	0.18	56,729	100.00

Table E.2. Scale Score Distribution—Mathematics Grade 4

SS	Freq.	%	Cum. Freq.	Cum. %
650	140	0.24	140	0.24
651	2	0.00	142	0.25
652	4	0.01	146	0.25
653	19	0.03	165	0.29
654	11	0.02	176	0.31
655	15	0.03	191	0.33
656	17	0.03	208	0.36
657	20	0.03	228	0.40
658	32	0.06	260	0.45
659	24	0.04	284	0.49
660	33	0.06	317	0.55
661	30	0.05	347	0.60
662	29	0.05	376	0.66
663	45	0.08	421	0.73
664	56	0.10	477	0.83
665	56	0.10	533	0.93
666	66	0.11	599	1.04
667	55	0.10	654	1.14
668	88	0.15	742	1.29
669	86	0.15	828	1.44
670	68	0.12	896	1.56
671	91	0.16	987	1.72
672	102	0.18	1,089	1.90
673	120	0.21	1,209	2.11
674	101	0.18	1,310	2.28
675	144	0.25	1,454	2.53
676	126	0.22	1,580	2.75
677	163	0.28	1,743	3.04
678	188	0.33	1,931	3.36
679	162	0.28	2,093	3.65
680	206	0.36	2,299	4.01
681	215	0.37	2,514	4.38
682	229	0.40	2,743	4.78
683	298	0.52	3,041	5.30
684	306	0.53	3,347	5.83
685	313	0.55	3,660	6.38
686	352	0.61	4,012	6.99
687	347	0.60	4,359	7.59
688	324	0.56	4,683	8.16
689	434	0.76	5,117	8.92
690	386	0.67	5,503	9.59
691	454	0.79	5,957	10.38
692	428	0.75	6,385	11.12
693	483	0.84	6,868	11.97
694	475	0.83	7,343	12.79
695	472	0.82	7,815	13.62
696	510	0.89	8,325	14.50
697	508	0.89	8,833	15.39
698	560	0.98	9,393	16.37
699	562	0.98	9,955	17.34
700	555	0.97	10,510	18.31
701	544	0.95	11,054	19.26
702	563	0.98	11,617	20.24
703	545	0.95	12,162	21.19
704	540	0.94	12,702	22.13
705	560	0.98	13,262	23.11
706	585	1.02	13,847	24.13
707	541	0.94	14,388	25.07
708	582	1.01	14,970	26.08
709	541	0.94	15,511	27.03
710	582	1.01	16,093	28.04
711	579	1.01	16,672	29.05
712	547	0.95	17,219	30.00
713	528	0.92	17,747	30.92
714	500	0.87	18,247	31.79
715	533	0.93	18,780	32.72
716	548	0.95	19,328	33.68
717	565	0.98	19,893	34.66
718	539	0.94	20,432	35.60
719	582	1.01	21,014	36.61
720	511	0.89	21,525	37.50
721	511	0.89	22,036	38.39
722	537	0.94	22,573	39.33
723	558	0.97	23,131	40.30
724	533	0.93	23,664	41.23
725	553	0.96	24,217	42.19
726	543	0.95	24,760	43.14
727	575	1.00	25,335	44.14
728	565	0.98	25,900	45.13
729	567	0.99	26,467	46.11
730	511	0.89	26,978	47.00
731	573	1.00	27,551	48.00
732	573	1.00	28,124	49.00
733	586	1.02	28,710	50.02
734	535	0.93	29,245	50.95
735	611	1.06	29,856	52.02
736	563	0.98	30,419	53.00
737	550	0.96	30,969	53.96
738	588	1.02	31,557	54.98
739	554	0.97	32,111	55.95
740	566	0.99	32,677	56.93
741	567	0.99	33,244	57.92
742	598	1.04	33,842	58.96
743	552	0.96	34,394	59.93
744	574	1.00	34,968	60.93
745	595	1.04	35,563	61.96
746	597	1.04	36,160	63.00
747	543	0.95	36,703	63.95
748	544	0.95	37,247	64.90
749	591	1.03	37,838	65.93

SS	Freq.	%	Cum. Freq.	Cum. %
750	557	0.97	38,395	66.90
751	539	0.94	38,934	67.84
752	527	0.92	39,461	68.75
753	552	0.96	40,013	69.72
754	540	0.94	40,553	70.66
755	535	0.93	41,088	71.59
756	507	0.88	41,595	72.47
757	456	0.79	42,051	73.27
758	480	0.84	42,531	74.10
759	493	0.86	43,024	74.96
760	478	0.83	43,502	75.79
761	473	0.82	43,975	76.62
762	494	0.86	44,469	77.48
763	502	0.87	44,971	78.35
764	482	0.84	45,453	79.19
765	425	0.74	45,878	79.93
766	416	0.72	46,294	80.66
767	437	0.76	46,731	81.42
768	439	0.76	47,170	82.18
769	421	0.73	47,591	82.92
770	435	0.76	48,026	83.68
771	372	0.65	48,398	84.32
772	391	0.68	48,789	85.01
773	410	0.71	49,199	85.72
774	383	0.67	49,582	86.39
775	350	0.61	49,932	87.00
776	329	0.57	50,261	87.57
777	319	0.56	50,580	88.13
778	328	0.57	50,908	88.70
779	303	0.53	51,211	89.23
780	290	0.51	51,501	89.73
781	325	0.57	51,826	90.30
782	269	0.47	52,095	90.77
783	264	0.46	52,359	91.23
784	277	0.48	52,636	91.71
785	245	0.43	52,881	92.14
786	253	0.44	53,134	92.58
787	246	0.43	53,380	93.00
788	234	0.41	53,614	93.41
789	182	0.32	53,796	93.73
790	183	0.32	53,979	94.05
791	211	0.37	54,190	94.42
792	215	0.37	54,405	94.79
793	168	0.29	54,573	95.08
794	157	0.27	54,730	95.36
795	164	0.29	54,894	95.64
796	138	0.24	55,032	95.88
797	154	0.27	55,186	96.15
798	156	0.27	55,342	96.42
799	134	0.23	55,476	96.66
800	131	0.23	55,607	96.88
801	134	0.23	55,741	97.12
802	118	0.21	55,859	97.32
803	99	0.17	55,958	97.50
804	101	0.18	56,059	97.67
805	74	0.13	56,133	97.80
806	66	0.11	56,199	97.92
807	83	0.14	56,282	98.06
808	72	0.13	56,354	98.19
809	72	0.13	56,426	98.31
810	59	0.10	56,485	98.41
811	58	0.10	56,543	98.52
812	49	0.09	56,592	98.60
813	56	0.10	56,648	98.70
814	40	0.07	56,688	98.77
815	57	0.10	56,745	98.87
816	60	0.10	56,805	98.97
817	42	0.07	56,847	99.05
818	37	0.06	56,884	99.11
819	33	0.06	56,917	99.17
820	36	0.06	56,953	99.23
821	17	0.03	56,970	99.26
822	25	0.04	56,995	99.30
823	34	0.06	57,029	99.36
824	35	0.06	57,064	99.42
825	21	0.04	57,085	99.46
826	25	0.04	57,110	99.50
827	23	0.04	57,133	99.54
828	12	0.02	57,145	99.56
829	12	0.02	57,157	99.59
830	6	0.01	57,163	99.60
831	5	0.01	57,168	99.60
832	6	0.01	57,174	99.61
833	24	0.04	57,198	99.66
834	6	0.01	57,204	99.67
835	30	0.05	57,234	99.72
836	14	0.02	57,248	99.74
837	22	0.04	57,270	99.78
838	7	0.01	57,277	99.79
839	15	0.03	57,292	99.82
840	5	0.01	57,297	99.83
842	7	0.01	57,304	99.84
844	2	0.00	57,306	99.84
846	1	0.00	57,307	99.85
847	2	0.00	57,309	99.85
849	4	0.01	57,313	99.86
850	82	0.14	57,395	100.00

Table E.3. Scale Score Distribution—Mathematics Grade 5

SS	Freq.	%	Cum. Freq.	Cum. %
650	129	0.23	129	0.23
653	2	0.00	131	0.23
654	9	0.02	140	0.25
655	2	0.00	142	0.25
656	3	0.01	145	0.26
657	3	0.01	148	0.26
658	31	0.05	179	0.32
659	19	0.03	198	0.35
660	20	0.04	218	0.39
661	17	0.03	235	0.42
662	18	0.03	253	0.45
663	48	0.08	301	0.53
664	36	0.06	337	0.60
665	45	0.08	382	0.67
666	12	0.02	394	0.70
667	37	0.07	431	0.76
668	20	0.04	451	0.80
669	40	0.07	491	0.87
670	28	0.05	519	0.92
671	51	0.09	570	1.01
672	67	0.12	637	1.13
673	57	0.10	694	1.23
674	86	0.15	780	1.38
675	88	0.16	868	1.53
676	84	0.15	952	1.68
677	78	0.14	1,030	1.82
678	70	0.12	1,100	1.94
679	127	0.22	1,227	2.17
680	160	0.28	1,387	2.45
681	165	0.29	1,552	2.74
682	150	0.27	1,702	3.01
683	172	0.30	1,874	3.31
684	175	0.31	2,049	3.62
685	171	0.30	2,220	3.92
686	230	0.41	2,450	4.33
687	237	0.42	2,687	4.75
688	255	0.45	2,942	5.20
689	272	0.48	3,214	5.68
690	284	0.50	3,498	6.18
691	336	0.59	3,834	6.77
692	328	0.58	4,162	7.35
693	346	0.61	4,508	7.96
694	386	0.68	4,894	8.65
695	401	0.71	5,295	9.35
696	422	0.75	5,717	10.10
697	448	0.79	6,165	10.89
698	459	0.81	6,624	11.70
699	518	0.92	7,142	12.62
700	465	0.82	7,607	13.44
701	493	0.87	8,100	14.31
702	513	0.91	8,613	15.22
703	507	0.90	9,120	16.11
704	549	0.97	9,669	17.08
705	521	0.92	10,190	18.00
706	556	0.98	10,746	18.99
707	559	0.99	11,305	19.97
708	510	0.90	11,815	20.87
709	550	0.97	12,365	21.85
710	530	0.94	12,895	22.78
711	569	1.01	13,464	23.79
712	561	0.99	14,025	24.78
713	532	0.94	14,557	25.72
714	558	0.99	15,115	26.70
715	575	1.02	15,690	27.72
716	550	0.97	16,240	28.69
717	583	1.03	16,823	29.72
718	563	0.99	17,386	30.72
719	600	1.06	17,986	31.78
720	602	1.06	18,588	32.84
721	591	1.04	19,179	33.88
722	563	0.99	19,742	34.88
723	546	0.96	20,288	35.84
724	573	1.01	20,861	36.86
725	612	1.08	21,473	37.94
726	599	1.06	22,072	39.00
727	563	0.99	22,635	39.99
728	595	1.05	23,230	41.04
729	546	0.96	23,776	42.01
730	583	1.03	24,359	43.04
731	571	1.01	24,930	44.05
732	557	0.98	25,487	45.03
733	596	1.05	26,083	46.08
734	593	1.05	26,676	47.13
735	588	1.04	27,264	48.17
736	596	1.05	27,860	49.22
737	583	1.03	28,443	50.25
738	620	1.10	29,063	51.35
739	581	1.03	29,644	52.37
740	580	1.02	30,224	53.40
741	629	1.11	30,853	54.51
742	594	1.05	31,447	55.56
743	546	0.96	31,993	56.52
744	613	1.08	32,606	57.61
745	617	1.09	33,223	58.70
746	597	1.05	33,820	59.75
747	572	1.01	34,392	60.76
748	563	0.99	34,955	61.76
749	523	0.92	35,478	62.68
750	559	0.99	36,037	63.67
751	526	0.93	36,563	64.60

SS	Freq.	%	Cum. Freq.	Cum. %
752	537	0.95	37,100	65.55
753	545	0.96	37,645	66.51
754	547	0.97	38,192	67.48
755	553	0.98	38,745	68.45
756	516	0.91	39,261	69.36
757	524	0.93	39,785	70.29
758	502	0.89	40,287	71.18
759	532	0.94	40,819	72.12
760	520	0.92	41,339	73.04
761	501	0.89	41,840	73.92
762	501	0.89	42,341	74.81
763	518	0.92	42,859	75.72
764	479	0.85	43,338	76.57
765	464	0.82	43,802	77.39
766	462	0.82	44,264	78.20
767	445	0.79	44,709	78.99
768	468	0.83	45,177	79.82
769	449	0.79	45,626	80.61
770	450	0.80	46,076	81.40
771	410	0.72	46,486	82.13
772	457	0.81	46,943	82.94
773	375	0.66	47,318	83.60
774	372	0.66	47,690	84.26
775	432	0.76	48,122	85.02
776	403	0.71	48,525	85.73
777	404	0.71	48,929	86.45
778	353	0.62	49,282	87.07
779	392	0.69	49,674	87.76
780	348	0.61	50,022	88.38
781	361	0.64	50,383	89.01
782	331	0.58	50,714	89.60
783	310	0.55	51,024	90.15
784	284	0.50	51,308	90.65
785	286	0.51	51,594	91.15
786	267	0.47	51,861	91.63
787	250	0.44	52,111	92.07
788	269	0.48	52,380	92.54
789	250	0.44	52,630	92.98
790	242	0.43	52,872	93.41
791	199	0.35	53,071	93.76
792	212	0.37	53,283	94.14
793	205	0.36	53,488	94.50
794	195	0.34	53,683	94.84
795	192	0.34	53,875	95.18
796	185	0.33	54,060	95.51
797	158	0.28	54,218	95.79
798	171	0.30	54,389	96.09
799	144	0.25	54,533	96.35
800	109	0.19	54,642	96.54
801	141	0.25	54,783	96.79
802	109	0.19	54,892	96.98
803	101	0.18	54,993	97.16
804	105	0.19	55,098	97.34
805	95	0.17	55,193	97.51
806	108	0.19	55,301	97.70
807	89	0.16	55,390	97.86
808	76	0.13	55,466	97.99
809	64	0.11	55,530	98.11
810	67	0.12	55,597	98.23
811	64	0.11	55,661	98.34
812	69	0.12	55,730	98.46
813	49	0.09	55,779	98.55
814	61	0.11	55,840	98.66
815	64	0.11	55,904	98.77
816	42	0.07	55,946	98.84
817	46	0.08	55,992	98.92
818	35	0.06	56,027	98.99
819	40	0.07	56,067	99.06
820	25	0.04	56,092	99.10
821	32	0.06	56,124	99.16
822	24	0.04	56,148	99.20
823	24	0.04	56,172	99.24
824	28	0.05	56,200	99.29
825	29	0.05	56,229	99.34
826	25	0.04	56,254	99.39
827	19	0.03	56,273	99.42
828	11	0.02	56,284	99.44
829	19	0.03	56,303	99.47
830	23	0.04	56,326	99.51
831	13	0.02	56,339	99.54
832	24	0.04	56,363	99.58
833	17	0.03	56,380	99.61
834	15	0.03	56,395	99.64
835	13	0.02	56,408	99.66
836	14	0.02	56,422	99.68
837	12	0.02	56,434	99.70
838	14	0.02	56,448	99.73
839	23	0.04	56,471	99.77
840	8	0.01	56,479	99.78
841	1	0.00	56,480	99.79
842	11	0.02	56,491	99.81
843	8	0.01	56,499	99.82
844	3	0.01	56,502	99.83
845	9	0.02	56,511	99.84
846	2	0.00	56,513	99.84
847	1	0.00	56,514	99.85
848	2	0.00	56,516	99.85
849	3	0.01	56,519	99.86
850	82	0.14	56,601	100.00

Table E.4. Scale Score Distribution—Mathematics Grade 6

SS	Freq.	%	Cum. Freq.	Cum. %
650	245	0.44	245	0.44
652	5	0.01	250	0.45
653	1	0.00	251	0.46
655	30	0.05	281	0.51
656	30	0.05	311	0.56
657	5	0.01	316	0.57
658	30	0.05	346	0.63
659	67	0.12	413	0.75
660	27	0.05	440	0.80
661	52	0.09	492	0.89
662	1	0.00	493	0.89
663	25	0.05	518	0.94
664	101	0.18	619	1.12
665	63	0.11	682	1.24
666	42	0.08	724	1.31
667	18	0.03	742	1.35
668	26	0.05	768	1.39
669	50	0.09	818	1.48
670	46	0.08	864	1.57
671	89	0.16	953	1.73
672	124	0.22	1,077	1.95
673	116	0.21	1,193	2.16
674	71	0.13	1,264	2.29
675	108	0.20	1,372	2.49
676	137	0.25	1,509	2.74
677	100	0.18	1,609	2.92
678	75	0.14	1,684	3.05
679	133	0.24	1,817	3.30
680	195	0.35	2,012	3.65
681	224	0.41	2,236	4.06
682	140	0.25	2,376	4.31
683	199	0.36	2,575	4.67
684	261	0.47	2,836	5.14
685	190	0.34	3,026	5.49
686	284	0.52	3,310	6.00
687	242	0.44	3,552	6.44
688	313	0.57	3,865	7.01
689	291	0.53	4,156	7.54
690	334	0.61	4,490	8.15
691	341	0.62	4,831	8.76
692	379	0.69	5,210	9.45
693	454	0.82	5,664	10.28
694	430	0.78	6,094	11.06
695	492	0.89	6,586	11.95
696	482	0.87	7,068	12.82
697	477	0.87	7,545	13.69
698	507	0.92	8,052	14.61
699	530	0.96	8,582	15.57
700	548	0.99	9,130	16.56
701	535	0.97	9,665	17.53
702	558	1.01	10,223	18.55
703	618	1.12	10,841	19.67
704	601	1.09	11,442	20.76
705	630	1.14	12,072	21.90
706	671	1.22	12,743	23.12
707	688	1.25	13,431	24.37
708	631	1.14	14,062	25.51
709	640	1.16	14,702	26.67
710	653	1.18	15,355	27.86
711	633	1.15	15,988	29.00
712	608	1.10	16,596	30.11
713	650	1.18	17,246	31.29
714	684	1.24	17,930	32.53
715	625	1.13	18,555	33.66
716	649	1.18	19,204	34.84
717	608	1.10	19,812	35.94
718	599	1.09	20,411	37.03
719	619	1.12	21,030	38.15
720	601	1.09	21,631	39.24
721	622	1.13	22,253	40.37
722	660	1.20	22,913	41.57
723	594	1.08	23,507	42.64
724	662	1.20	24,169	43.84
725	645	1.17	24,814	45.01
726	571	1.04	25,385	46.05
727	660	1.20	26,045	47.25
728	610	1.11	26,655	48.35
729	652	1.18	27,307	49.54
730	628	1.14	27,935	50.68
731	616	1.12	28,551	51.79
732	641	1.16	29,192	52.96
733	618	1.12	29,810	54.08
734	587	1.06	30,397	55.14
735	609	1.10	31,006	56.25
736	575	1.04	31,581	57.29
737	576	1.04	32,157	58.34
738	589	1.07	32,746	59.40
739	587	1.06	33,333	60.47
740	557	1.01	33,890	61.48
741	596	1.08	34,486	62.56
742	566	1.03	35,052	63.59
743	602	1.09	35,654	64.68
744	557	1.01	36,211	65.69
745	561	1.02	36,772	66.71
746	587	1.06	37,359	67.77
747	556	1.01	37,915	68.78
748	540	0.98	38,455	69.76
749	572	1.04	39,027	70.80
750	578	1.05	39,605	71.85
751	528	0.96	40,133	72.80

SS	Freq.	%	Cum. Freq.	Cum. %
752	504	0.91	40,637	73.72
753	524	0.95	41,161	74.67
754	511	0.93	41,672	75.60
755	518	0.94	42,190	76.54
756	493	0.89	42,683	77.43
757	476	0.86	43,159	78.29
758	467	0.85	43,626	79.14
759	455	0.83	44,081	79.97
760	478	0.87	44,559	80.83
761	464	0.84	45,023	81.68
762	428	0.78	45,451	82.45
763	442	0.80	45,893	83.25
764	407	0.74	46,300	83.99
765	416	0.75	46,716	84.75
766	366	0.66	47,082	85.41
767	389	0.71	47,471	86.12
768	377	0.68	47,848	86.80
769	364	0.66	48,212	87.46
770	353	0.64	48,565	88.10
771	342	0.62	48,907	88.72
772	346	0.63	49,253	89.35
773	329	0.60	49,582	89.95
774	274	0.50	49,856	90.44
775	282	0.51	50,138	90.95
776	266	0.48	50,404	91.44
777	248	0.45	50,652	91.89
778	261	0.47	50,913	92.36
779	244	0.44	51,157	92.80
780	250	0.45	51,407	93.26
781	223	0.40	51,630	93.66
782	227	0.41	51,857	94.07
783	190	0.34	52,047	94.42
784	193	0.35	52,240	94.77
785	187	0.34	52,427	95.11
786	172	0.31	52,599	95.42
787	168	0.30	52,767	95.72
788	151	0.27	52,918	96.00
789	157	0.28	53,075	96.28
790	122	0.22	53,197	96.50
791	127	0.23	53,324	96.73
792	125	0.23	53,449	96.96
793	118	0.21	53,567	97.18
794	105	0.19	53,672	97.37
795	115	0.21	53,787	97.57
796	91	0.17	53,878	97.74
797	108	0.20	53,986	97.94
798	90	0.16	54,076	98.10
799	78	0.14	54,154	98.24
800	68	0.12	54,222	98.36
801	59	0.11	54,281	98.47
802	58	0.11	54,339	98.58
803	71	0.13	54,410	98.70
804	51	0.09	54,461	98.80
805	51	0.09	54,512	98.89
806	42	0.08	54,554	98.97
807	45	0.08	54,599	99.05
808	44	0.08	54,643	99.13
809	42	0.08	54,685	99.20
810	45	0.08	54,730	99.29
811	25	0.05	54,755	99.33
812	23	0.04	54,778	99.37
813	24	0.04	54,802	99.42
814	22	0.04	54,824	99.46
815	21	0.04	54,845	99.49
816	19	0.03	54,864	99.53
817	19	0.03	54,883	99.56
818	14	0.03	54,897	99.59
819	19	0.03	54,916	99.62
820	21	0.04	54,937	99.66
821	8	0.01	54,945	99.68
822	14	0.03	54,959	99.70
823	10	0.02	54,969	99.72
824	9	0.02	54,978	99.74
825	11	0.02	54,989	99.76
826	9	0.02	54,998	99.77
827	8	0.01	55,006	99.79
828	5	0.01	55,011	99.80
829	10	0.02	55,021	99.81
830	10	0.02	55,031	99.83
831	1	0.00	55,032	99.83
832	9	0.02	55,041	99.85
833	6	0.01	55,047	99.86
834	4	0.01	55,051	99.87
835	5	0.01	55,056	99.88
836	10	0.02	55,066	99.89
837	3	0.01	55,069	99.90
838	4	0.01	55,073	99.91
839	2	0.00	55,075	99.91
840	2	0.00	55,077	99.91
841	2	0.00	55,079	99.92
843	2	0.00	55,081	99.92
844	7	0.01	55,088	99.93
846	4	0.01	55,092	99.94
847	1	0.00	55,093	99.94
848	2	0.00	55,095	99.95
850	29	0.05	55,124	100.00

Table E.5. Scale Score Distribution—Mathematics Grade 7

SS	Freq.	%	Cum. Freq.	Cum. %
650	149	0.28	149	0.28
651	7	0.01	156	0.29
652	7	0.01	163	0.30
653	8	0.01	171	0.32
654	28	0.05	199	0.37
656	38	0.07	237	0.44
657	18	0.03	255	0.47
658	35	0.06	290	0.54
660	10	0.02	300	0.56
661	21	0.04	321	0.60
662	8	0.01	329	0.61
663	11	0.02	340	0.63
664	30	0.06	370	0.69
665	39	0.07	409	0.76
666	19	0.04	428	0.79
667	33	0.06	461	0.86
668	45	0.08	506	0.94
669	29	0.05	535	0.99
670	36	0.07	571	1.06
671	49	0.09	620	1.15
672	60	0.11	680	1.26
673	49	0.09	729	1.35
674	57	0.11	786	1.46
675	55	0.10	841	1.56
676	112	0.21	953	1.77
677	56	0.10	1,009	1.87
678	87	0.16	1,096	2.03
679	71	0.13	1,167	2.17
680	103	0.19	1,270	2.36
681	132	0.24	1,402	2.60
682	147	0.27	1,549	2.87
683	113	0.21	1,662	3.08
684	140	0.26	1,802	3.34
685	191	0.35	1,993	3.70
686	185	0.34	2,178	4.04
687	192	0.36	2,370	4.40
688	200	0.37	2,570	4.77
689	199	0.37	2,769	5.14
690	290	0.54	3,059	5.68
691	244	0.45	3,303	6.13
692	273	0.51	3,576	6.63
693	329	0.61	3,905	7.24
694	338	0.63	4,243	7.87
695	371	0.69	4,614	8.56
696	398	0.74	5,012	9.30
697	431	0.80	5,443	10.10
698	441	0.82	5,884	10.92
699	465	0.86	6,349	11.78
700	479	0.89	6,828	12.67
701	537	1.00	7,365	13.66
702	593	1.10	7,958	14.76
703	569	1.06	8,527	15.82
704	568	1.05	9,095	16.87
705	571	1.06	9,666	17.93
706	616	1.14	10,282	19.08
707	609	1.13	10,891	20.21
708	614	1.14	11,505	21.34
709	638	1.18	12,143	22.53
710	670	1.24	12,813	23.77
711	624	1.16	13,437	24.93
712	662	1.23	14,099	26.16
713	675	1.25	14,774	27.41
714	699	1.30	15,473	28.71
715	617	1.14	16,090	29.85
716	671	1.24	16,761	31.10
717	657	1.22	17,418	32.31
718	702	1.30	18,120	33.62
719	665	1.23	18,785	34.85
720	644	1.19	19,429	36.05
721	668	1.24	20,097	37.28
722	675	1.25	20,772	38.54
723	684	1.27	21,456	39.81
724	635	1.18	22,091	40.98
725	625	1.16	22,716	42.14
726	644	1.19	23,360	43.34
727	635	1.18	23,995	44.52
728	659	1.22	24,654	45.74
729	670	1.24	25,324	46.98
730	648	1.20	25,972	48.18
731	633	1.17	26,605	49.36
732	655	1.22	27,260	50.57
733	684	1.27	27,944	51.84
734	641	1.19	28,585	53.03
735	642	1.19	29,227	54.22
736	666	1.24	29,893	55.46
737	631	1.17	30,524	56.63
738	603	1.12	31,127	57.75
739	632	1.17	31,759	58.92
740	599	1.11	32,358	60.03
741	623	1.16	32,981	61.19
742	625	1.16	33,606	62.35
743	620	1.15	34,226	63.50
744	576	1.07	34,802	64.57
745	590	1.09	35,392	65.66
746	653	1.21	36,045	66.87
747	623	1.16	36,668	68.03
748	615	1.14	37,283	69.17
749	565	1.05	37,848	70.22
750	575	1.07	38,423	71.28
751	555	1.03	38,978	72.31

SS	Freq.	%	Cum. Freq.	Cum. %
752	558	1.04	39,536	73.35
753	594	1.10	40,130	74.45
754	560	1.04	40,690	75.49
755	563	1.04	41,253	76.53
756	493	0.91	41,746	77.45
757	532	0.99	42,278	78.43
758	467	0.87	42,745	79.30
759	507	0.94	43,252	80.24
760	504	0.94	43,756	81.18
761	476	0.88	44,232	82.06
762	478	0.89	44,710	82.95
763	460	0.85	45,170	83.80
764	449	0.83	45,619	84.63
765	422	0.78	46,041	85.42
766	407	0.76	46,448	86.17
767	423	0.78	46,871	86.96
768	371	0.69	47,242	87.64
769	396	0.73	47,638	88.38
770	347	0.64	47,985	89.02
771	323	0.60	48,308	89.62
772	327	0.61	48,635	90.23
773	337	0.63	48,972	90.85
774	293	0.54	49,265	91.40
775	312	0.58	49,577	91.98
776	269	0.50	49,846	92.48
777	248	0.46	50,094	92.94
778	242	0.45	50,336	93.38
779	216	0.40	50,552	93.79
780	211	0.39	50,763	94.18
781	225	0.42	50,988	94.59
782	197	0.37	51,185	94.96
783	168	0.31	51,353	95.27
784	182	0.34	51,535	95.61
785	159	0.29	51,694	95.90
786	148	0.27	51,842	96.18
787	135	0.25	51,977	96.43
788	148	0.27	52,125	96.70
789	128	0.24	52,253	96.94
790	100	0.19	52,353	97.13
791	112	0.21	52,465	97.33
792	103	0.19	52,568	97.53
793	92	0.17	52,660	97.70
794	98	0.18	52,758	97.88
795	96	0.18	52,854	98.06
796	80	0.15	52,934	98.20
797	76	0.14	53,010	98.35
798	69	0.13	53,079	98.47
799	76	0.14	53,155	98.61
800	64	0.12	53,219	98.73
801	54	0.10	53,273	98.83
802	40	0.07	53,313	98.91
803	53	0.10	53,366	99.01
804	32	0.06	53,398	99.06
805	34	0.06	53,432	99.13
806	36	0.07	53,468	99.19
807	33	0.06	53,501	99.26
808	36	0.07	53,537	99.32
809	25	0.05	53,562	99.37
810	21	0.04	53,583	99.41
811	22	0.04	53,605	99.45
812	25	0.05	53,630	99.50
813	28	0.05	53,658	99.55
814	19	0.04	53,677	99.58
815	17	0.03	53,694	99.61
816	14	0.03	53,708	99.64
817	17	0.03	53,725	99.67
818	11	0.02	53,736	99.69
819	9	0.02	53,745	99.71
820	11	0.02	53,756	99.73
821	10	0.02	53,766	99.75
822	10	0.02	53,776	99.77
823	6	0.01	53,782	99.78
824	5	0.01	53,787	99.79
825	10	0.02	53,797	99.81
826	5	0.01	53,802	99.81
827	4	0.01	53,806	99.82
828	6	0.01	53,812	99.83
829	6	0.01	53,818	99.84
830	4	0.01	53,822	99.85
831	6	0.01	53,828	99.86
832	4	0.01	53,832	99.87
833	4	0.01	53,836	99.88
834	4	0.01	53,840	99.88
835	4	0.01	53,844	99.89
836	2	0.00	53,846	99.90
837	7	0.01	53,853	99.91
838	1	0.00	53,854	99.91
839	3	0.01	53,857	99.92
840	1	0.00	53,858	99.92
841	1	0.00	53,859	99.92
842	2	0.00	53,861	99.92
843	3	0.01	53,864	99.93
844	1	0.00	53,865	99.93
845	2	0.00	53,867	99.94
848	2	0.00	53,869	99.94
850	33	0.06	53,902	100.00

Table E.6. Scale Score Distribution—Mathematics Grade 8

SS	Freq.	%	Cum. Freq.	Cum. %
650	609	1.20	609	1.20
651	43	0.08	652	1.28
652	24	0.05	676	1.33
653	43	0.08	719	1.42
654	42	0.08	761	1.50
655	41	0.08	802	1.58
656	74	0.15	876	1.72
657	62	0.12	938	1.85
658	76	0.15	1,014	2.00
659	95	0.19	1,109	2.18
660	68	0.13	1,177	2.32
661	68	0.13	1,245	2.45
662	83	0.16	1,328	2.61
663	75	0.15	1,403	2.76
664	75	0.15	1,478	2.91
665	103	0.20	1,581	3.11
666	134	0.26	1,715	3.38
667	156	0.31	1,871	3.68
668	138	0.27	2,009	3.95
669	165	0.32	2,174	4.28
670	158	0.31	2,332	4.59
671	163	0.32	2,495	4.91
672	155	0.31	2,650	5.22
673	191	0.38	2,841	5.59
674	198	0.39	3,039	5.98
675	215	0.42	3,254	6.41
676	215	0.42	3,469	6.83
677	248	0.49	3,717	7.32
678	246	0.48	3,963	7.80
679	262	0.52	4,225	8.32
680	314	0.62	4,539	8.93
681	315	0.62	4,854	9.55
682	332	0.65	5,186	10.21
683	331	0.65	5,517	10.86
684	321	0.63	5,838	11.49
685	338	0.67	6,176	12.16
686	412	0.81	6,588	12.97
687	391	0.77	6,979	13.74
688	379	0.75	7,358	14.48
689	399	0.79	7,757	15.27
690	418	0.82	8,175	16.09
691	397	0.78	8,572	16.87
692	477	0.94	9,049	17.81
693	509	1.00	9,558	18.81
694	503	0.99	10,061	19.80
695	449	0.88	10,510	20.69
696	494	0.97	11,004	21.66
697	488	0.96	11,492	22.62
698	513	1.01	12,005	23.63
699	490	0.96	12,495	24.60
700	541	1.06	13,036	25.66
701	537	1.06	13,573	26.72
702	538	1.06	14,111	27.78
703	517	1.02	14,628	28.79
704	518	1.02	15,146	29.81
705	483	0.95	15,629	30.76
706	481	0.95	16,110	31.71
707	523	1.03	16,633	32.74
708	509	1.00	17,142	33.74
709	487	0.96	17,629	34.70
710	468	0.92	18,097	35.62
711	494	0.97	18,591	36.59
712	488	0.96	19,079	37.55
713	496	0.98	19,575	38.53
714	467	0.92	20,042	39.45
715	461	0.91	20,503	40.36
716	491	0.97	20,994	41.32
717	454	0.89	21,448	42.22
718	502	0.99	21,950	43.21
719	477	0.94	22,427	44.15
720	487	0.96	22,914	45.10
721	444	0.87	23,358	45.98
722	415	0.82	23,773	46.79
723	432	0.85	24,205	47.64
724	400	0.79	24,605	48.43
725	431	0.85	25,036	49.28
726	431	0.85	25,467	50.13
727	391	0.77	25,858	50.90
728	445	0.88	26,303	51.77
729	421	0.83	26,724	52.60
730	411	0.81	27,135	53.41
731	380	0.75	27,515	54.16
732	414	0.81	27,929	54.98
733	397	0.78	28,326	55.76
734	365	0.72	28,691	56.48
735	384	0.76	29,075	57.23
736	374	0.74	29,449	57.97
737	382	0.75	29,831	58.72
738	377	0.74	30,208	59.46
739	375	0.74	30,583	60.20
740	384	0.76	30,967	60.96
741	392	0.77	31,359	61.73
742	367	0.72	31,726	62.45
743	380	0.75	32,106	63.20
744	365	0.72	32,471	63.92
745	349	0.69	32,820	64.60
746	351	0.69	33,171	65.29
747	377	0.74	33,548	66.04
748	358	0.70	33,906	66.74
749	376	0.74	34,282	67.48

SS	Freq.	%	Cum. Freq.	Cum. %
750	387	0.76	34,669	68.24
751	399	0.79	35,068	69.03
752	381	0.75	35,449	69.78
753	344	0.68	35,793	70.45
754	394	0.78	36,187	71.23
755	351	0.69	36,538	71.92
756	346	0.68	36,884	72.60
757	339	0.67	37,223	73.27
758	330	0.65	37,553	73.92
759	284	0.56	37,837	74.48
760	322	0.63	38,159	75.11
761	317	0.62	38,476	75.74
762	295	0.58	38,771	76.32
763	355	0.70	39,126	77.02
764	272	0.54	39,398	77.55
765	291	0.57	39,689	78.12
766	294	0.58	39,983	78.70
767	317	0.62	40,300	79.33
768	291	0.57	40,591	79.90
769	284	0.56	40,875	80.46
770	297	0.58	41,172	81.04
771	255	0.50	41,427	81.54
772	300	0.59	41,727	82.13
773	272	0.54	41,999	82.67
774	270	0.53	42,269	83.20
775	252	0.50	42,521	83.70
776	272	0.54	42,793	84.23
777	244	0.48	43,037	84.71
778	238	0.47	43,275	85.18
779	212	0.42	43,487	85.60
780	242	0.48	43,729	86.08
781	254	0.50	43,983	86.58
782	245	0.48	44,228	87.06
783	243	0.48	44,471	87.54
784	226	0.44	44,697	87.98
785	208	0.41	44,905	88.39
786	202	0.40	45,107	88.79
787	189	0.37	45,296	89.16
788	215	0.42	45,511	89.58
789	189	0.37	45,700	89.96
790	172	0.34	45,872	90.29
791	198	0.39	46,070	90.68
792	196	0.39	46,266	91.07
793	180	0.35	46,446	91.42
794	183	0.36	46,629	91.78
795	150	0.30	46,779	92.08
796	179	0.35	46,958	92.43
797	184	0.36	47,142	92.79
798	156	0.31	47,298	93.10
799	159	0.31	47,457	93.41
800	128	0.25	47,585	93.67
801	135	0.27	47,720	93.93
802	138	0.27	47,858	94.20
803	158	0.31	48,016	94.51
804	140	0.28	48,156	94.79
805	129	0.25	48,285	95.04
806	129	0.25	48,414	95.30
807	129	0.25	48,543	95.55
808	122	0.24	48,665	95.79
809	122	0.24	48,787	96.03
810	116	0.23	48,903	96.26
811	105	0.21	49,008	96.47
812	97	0.19	49,105	96.66
813	114	0.22	49,219	96.88
814	108	0.21	49,327	97.09
815	74	0.15	49,401	97.24
816	74	0.15	49,475	97.39
817	74	0.15	49,549	97.53
818	70	0.14	49,619	97.67
819	89	0.18	49,708	97.84
820	67	0.13	49,775	97.98
821	61	0.12	49,836	98.10
822	55	0.11	49,891	98.20
823	40	0.08	49,931	98.28
824	47	0.09	49,978	98.38
825	55	0.11	50,033	98.48
826	43	0.08	50,076	98.57
827	37	0.07	50,113	98.64
828	33	0.06	50,146	98.71
829	35	0.07	50,181	98.78
830	27	0.05	50,208	98.83
831	26	0.05	50,234	98.88
832	33	0.06	50,267	98.94
833	23	0.05	50,290	98.99
834	25	0.05	50,315	99.04
835	35	0.07	50,350	99.11
836	16	0.03	50,366	99.14
837	23	0.05	50,389	99.19
838	30	0.06	50,419	99.24
839	25	0.05	50,444	99.29
840	22	0.04	50,466	99.34
841	16	0.03	50,482	99.37
842	14	0.03	50,496	99.40
843	17	0.03	50,513	99.43
844	12	0.02	50,525	99.45
845	15	0.03	50,540	99.48
846	18	0.04	50,558	99.52
847	10	0.02	50,568	99.54
848	9	0.02	50,577	99.56
849	5	0.01	50,582	99.56
850	221	0.44	50,803	100.00

Table E.7. Scale Score Distribution—ELA Grade 3

SS	Freq.	%	Cum. Freq.	Cum. %
650	1,347	2.47	1,347	2.47
651	81	0.15	1,428	2.62
652	96	0.18	1,524	2.79
653	113	0.21	1,637	3.00
654	105	0.19	1,742	3.19
655	126	0.23	1,868	3.42
656	117	0.21	1,985	3.64
657	97	0.18	2,082	3.81
658	136	0.25	2,218	4.06
659	126	0.23	2,344	4.29
660	138	0.25	2,482	4.55
661	131	0.24	2,613	4.79
662	173	0.32	2,786	5.10
663	150	0.27	2,936	5.38
664	181	0.33	3,117	5.71
665	172	0.32	3,289	6.03
666	187	0.34	3,476	6.37
667	173	0.32	3,649	6.69
668	203	0.37	3,852	7.06
669	180	0.33	4,032	7.39
670	211	0.39	4,243	7.77
671	231	0.42	4,474	8.20
672	231	0.42	4,705	8.62
673	257	0.47	4,962	9.09
674	248	0.45	5,210	9.55
675	268	0.49	5,478	10.04
676	260	0.48	5,738	10.51
677	256	0.47	5,994	10.98
678	250	0.46	6,244	11.44
679	259	0.47	6,503	11.92
680	233	0.43	6,736	12.34
681	278	0.51	7,014	12.85
682	248	0.45	7,262	13.31
683	260	0.48	7,522	13.78
684	255	0.47	7,777	14.25
685	272	0.50	8,049	14.75
686	306	0.56	8,355	15.31
687	262	0.48	8,617	15.79
688	254	0.47	8,871	16.25
689	268	0.49	9,139	16.75
690	311	0.57	9,450	17.31
691	274	0.50	9,724	17.82
692	253	0.46	9,977	18.28
693	291	0.53	10,268	18.81
694	279	0.51	10,547	19.32
695	320	0.59	10,867	19.91
696	267	0.49	11,134	20.40
697	304	0.56	11,438	20.96
698	275	0.50	11,713	21.46
699	302	0.55	12,015	22.01
700	309	0.57	12,324	22.58
701	289	0.53	12,613	23.11
702	300	0.55	12,913	23.66
703	277	0.51	13,190	24.17
704	307	0.56	13,497	24.73
705	306	0.56	13,803	25.29
706	298	0.55	14,101	25.84
707	321	0.59	14,422	26.43
708	355	0.65	14,777	27.08
709	311	0.57	15,088	27.65
710	335	0.61	15,423	28.26
711	341	0.62	15,764	28.88
712	344	0.63	16,108	29.51
713	351	0.64	16,459	30.16
714	326	0.60	16,785	30.75
715	348	0.64	17,133	31.39
716	377	0.69	17,510	32.08
717	357	0.65	17,867	32.74
718	388	0.71	18,255	33.45
719	375	0.69	18,630	34.14
720	382	0.70	19,012	34.84
721	369	0.68	19,381	35.51
722	343	0.63	19,724	36.14
723	401	0.73	20,125	36.87
724	377	0.69	20,502	37.57
725	386	0.71	20,888	38.27
726	389	0.71	21,277	38.99
727	440	0.81	21,717	39.79
728	393	0.72	22,110	40.51
729	436	0.80	22,546	41.31
730	388	0.71	22,934	42.02
731	464	0.85	23,398	42.87
732	387	0.71	23,785	43.58
733	429	0.79	24,214	44.37
734	452	0.83	24,666	45.19
735	449	0.82	25,115	46.02
736	447	0.82	25,562	46.84
737	446	0.82	26,008	47.65
738	490	0.90	26,498	48.55
739	470	0.86	26,968	49.41
740	448	0.82	27,416	50.23
741	452	0.83	27,868	51.06
742	492	0.90	28,360	51.96
743	446	0.82	28,806	52.78
744	485	0.89	29,291	53.67
745	472	0.86	29,763	54.53
746	456	0.84	30,219	55.37
747	454	0.83	30,673	56.20
748	467	0.86	31,140	57.06
749	460	0.84	31,600	57.90

SS	Freq.	%	Cum. Freq.	Cum. %
750	464	0.85	32,064	58.75
751	514	0.94	32,578	59.69
752	486	0.89	33,064	60.58
753	477	0.87	33,541	61.46
754	479	0.88	34,020	62.33
755	477	0.87	34,497	63.21
756	424	0.78	34,921	63.98
757	461	0.84	35,382	64.83
758	461	0.84	35,843	65.67
759	473	0.87	36,316	66.54
760	451	0.83	36,767	67.37
761	451	0.83	37,218	68.19
762	433	0.79	37,651	68.99
763	495	0.91	38,146	69.89
764	453	0.83	38,599	70.72
765	466	0.85	39,065	71.58
766	441	0.81	39,506	72.39
767	421	0.77	39,927	73.16
768	423	0.78	40,350	73.93
769	420	0.77	40,770	74.70
770	429	0.79	41,199	75.49
771	393	0.72	41,592	76.21
772	425	0.78	42,017	76.99
773	402	0.74	42,419	77.72
774	403	0.74	42,822	78.46
775	400	0.73	43,222	79.19
776	386	0.71	43,608	79.90
777	361	0.66	43,969	80.56
778	438	0.80	44,407	81.37
779	355	0.65	44,762	82.02
780	373	0.68	45,135	82.70
781	351	0.64	45,486	83.34
782	354	0.65	45,840	83.99
783	352	0.64	46,192	84.64
784	334	0.61	46,526	85.25
785	300	0.55	46,826	85.80
786	308	0.56	47,134	86.36
787	322	0.59	47,456	86.95
788	311	0.57	47,767	87.52
789	285	0.52	48,052	88.04
790	276	0.51	48,328	88.55
791	289	0.53	48,617	89.08
792	279	0.51	48,896	89.59
793	264	0.48	49,160	90.07
794	253	0.46	49,413	90.54
795	254	0.47	49,667	91.00
796	229	0.42	49,896	91.42
797	218	0.40	50,114	91.82
798	192	0.35	50,306	92.17
799	202	0.37	50,508	92.54
800	190	0.35	50,698	92.89
801	199	0.36	50,897	93.26
802	176	0.32	51,073	93.58
803	178	0.33	51,251	93.91
804	175	0.32	51,426	94.23
805	176	0.32	51,602	94.55
806	157	0.29	51,759	94.84
807	152	0.28	51,911	95.12
808	160	0.29	52,071	95.41
809	126	0.23	52,197	95.64
810	165	0.30	52,362	95.94
811	92	0.17	52,454	96.11
812	135	0.25	52,589	96.36
813	82	0.15	52,671	96.51
814	97	0.18	52,768	96.69
815	119	0.22	52,887	96.90
816	70	0.13	52,957	97.03
817	114	0.21	53,071	97.24
818	67	0.12	53,138	97.36
819	106	0.19	53,244	97.56
820	71	0.13	53,315	97.69
821	70	0.13	53,385	97.82
822	59	0.11	53,444	97.92
823	78	0.14	53,522	98.07
824	69	0.13	53,591	98.19
825	82	0.15	53,673	98.34
826	37	0.07	53,710	98.41
827	57	0.10	53,767	98.52
828	48	0.09	53,815	98.60
829	39	0.07	53,854	98.68
830	61	0.11	53,915	98.79
831	26	0.05	53,941	98.83
832	33	0.06	53,974	98.90
833	45	0.08	54,019	98.98
834	48	0.09	54,067	99.07
835	29	0.05	54,096	99.12
836	37	0.07	54,133	99.19
837	22	0.04	54,155	99.23
838	20	0.04	54,175	99.26
839	35	0.06	54,210	99.33
840	18	0.03	54,228	99.36
841	33	0.06	54,261	99.42
842	24	0.04	54,285	99.46
843	14	0.03	54,299	99.49
844	13	0.02	54,312	99.51
845	22	0.04	54,334	99.55
846	17	0.03	54,351	99.59
847	17	0.03	54,368	99.62
848	14	0.03	54,382	99.64
849	10	0.02	54,392	99.66
850	185	0.34	54,577	100.00

Table E.8. Scale Score Distribution—ELA Grade 4

SS	Freq.	%	Cum. Freq.	Cum. %
650	305	0.55	305	0.55
651	22	0.04	327	0.59
652	21	0.04	348	0.62
653	35	0.06	383	0.69
654	22	0.04	405	0.73
655	37	0.07	442	0.79
656	29	0.05	471	0.85
657	23	0.04	494	0.89
658	36	0.06	530	0.95
659	38	0.07	568	1.02
660	42	0.08	610	1.09
661	42	0.08	652	1.17
662	46	0.08	698	1.25
663	55	0.10	753	1.35
664	71	0.13	824	1.48
665	64	0.11	888	1.59
666	59	0.11	947	1.70
667	66	0.12	1,013	1.82
668	86	0.15	1,099	1.97
669	105	0.19	1,204	2.16
670	100	0.18	1,304	2.34
671	105	0.19	1,409	2.53
672	134	0.24	1,543	2.77
673	116	0.21	1,659	2.98
674	138	0.25	1,797	3.23
675	123	0.22	1,920	3.45
676	133	0.24	2,053	3.69
677	137	0.25	2,190	3.93
678	186	0.33	2,376	4.26
679	183	0.33	2,559	4.59
680	184	0.33	2,743	4.92
681	212	0.38	2,955	5.30
682	186	0.33	3,141	5.64
683	201	0.36	3,342	6.00
684	214	0.38	3,556	6.38
685	221	0.40	3,777	6.78
686	246	0.44	4,023	7.22
687	277	0.50	4,300	7.72
688	262	0.47	4,562	8.19
689	283	0.51	4,845	8.70
690	268	0.48	5,113	9.18
691	297	0.53	5,410	9.71
692	273	0.49	5,683	10.20
693	324	0.58	6,007	10.78
694	306	0.55	6,313	11.33
695	322	0.58	6,635	11.91
696	369	0.66	7,004	12.57
697	353	0.63	7,357	13.21
698	371	0.67	7,728	13.87
699	362	0.65	8,090	14.52
700	358	0.64	8,448	15.16
701	328	0.59	8,776	15.75
702	353	0.63	9,129	16.39
703	395	0.71	9,524	17.10
704	406	0.73	9,930	17.82
705	399	0.72	10,329	18.54
706	377	0.68	10,706	19.22
707	409	0.73	11,115	19.95
708	377	0.68	11,492	20.63
709	398	0.71	11,890	21.34
710	440	0.79	12,330	22.13
711	377	0.68	12,707	22.81
712	412	0.74	13,119	23.55
713	458	0.82	13,577	24.37
714	436	0.78	14,013	25.15
715	430	0.77	14,443	25.93
716	487	0.87	14,930	26.80
717	468	0.84	15,398	27.64
718	418	0.75	15,816	28.39
719	486	0.87	16,302	29.26
720	481	0.86	16,783	30.13
721	465	0.83	17,248	30.96
722	483	0.87	17,731	31.83
723	461	0.83	18,192	32.65
724	471	0.85	18,663	33.50
725	495	0.89	19,158	34.39
726	499	0.90	19,657	35.28
727	475	0.85	20,132	36.14
728	498	0.89	20,630	37.03
729	490	0.88	21,120	37.91
730	539	0.97	21,659	38.88
731	568	1.02	22,227	39.90
732	502	0.90	22,729	40.80
733	508	0.91	23,237	41.71
734	539	0.97	23,776	42.68
735	568	1.02	24,344	43.70
736	582	1.04	24,926	44.74
737	579	1.04	25,505	45.78
738	573	1.03	26,078	46.81
739	572	1.03	26,650	47.84
740	596	1.07	27,246	48.91
741	566	1.02	27,812	49.92
742	561	1.01	28,373	50.93
743	559	1.00	28,932	51.93
744	577	1.04	29,509	52.97
745	533	0.96	30,042	53.93
746	573	1.03	30,615	54.95
747	554	0.99	31,169	55.95
748	579	1.04	31,748	56.99
749	571	1.02	32,319	58.01

SS	Freq.	%	Cum. Freq.	Cum. %
750	562	1.01	32,881	59.02
751	558	1.00	33,439	60.02
752	568	1.02	34,007	61.04
753	570	1.02	34,577	62.07
754	551	0.99	35,128	63.06
755	509	0.91	35,637	63.97
756	553	0.99	36,190	64.96
757	513	0.92	36,703	65.88
758	534	0.96	37,237	66.84
759	537	0.96	37,774	67.80
760	518	0.93	38,292	68.73
761	501	0.90	38,793	69.63
762	473	0.85	39,266	70.48
763	522	0.94	39,788	71.42
764	533	0.96	40,321	72.38
765	558	1.00	40,879	73.38
766	443	0.80	41,322	74.17
767	449	0.81	41,771	74.98
768	461	0.83	42,232	75.81
769	464	0.83	42,696	76.64
770	453	0.81	43,149	77.45
771	418	0.75	43,567	78.20
772	457	0.82	44,024	79.02
773	436	0.78	44,460	79.81
774	420	0.75	44,880	80.56
775	425	0.76	45,305	81.32
776	388	0.70	45,693	82.02
777	450	0.81	46,143	82.83
778	357	0.64	46,500	83.47
779	396	0.71	46,896	84.18
780	394	0.71	47,290	84.89
781	378	0.68	47,668	85.56
782	343	0.62	48,011	86.18
783	340	0.61	48,351	86.79
784	328	0.59	48,679	87.38
785	317	0.57	48,996	87.95
786	290	0.52	49,286	88.47
787	280	0.50	49,566	88.97
788	278	0.50	49,844	89.47
789	313	0.56	50,157	90.03
790	292	0.52	50,449	90.56
791	260	0.47	50,709	91.02
792	272	0.49	50,981	91.51
793	259	0.46	51,240	91.98
794	230	0.41	51,470	92.39
795	204	0.37	51,674	92.76
796	211	0.38	51,885	93.13
797	215	0.39	52,100	93.52
798	204	0.37	52,304	93.89
799	201	0.36	52,505	94.25
800	150	0.27	52,655	94.52
801	155	0.28	52,810	94.79
802	159	0.29	52,969	95.08
803	152	0.27	53,121	95.35
804	150	0.27	53,271	95.62
805	138	0.25	53,409	95.87
806	139	0.25	53,548	96.12
807	117	0.21	53,665	96.33
808	126	0.23	53,791	96.56
809	128	0.23	53,919	96.79
810	107	0.19	54,026	96.98
811	99	0.18	54,125	97.15
812	101	0.18	54,226	97.34
813	99	0.18	54,325	97.51
814	70	0.13	54,395	97.64
815	82	0.15	54,477	97.79
816	59	0.11	54,536	97.89
817	72	0.13	54,608	98.02
818	66	0.12	54,674	98.14
819	67	0.12	54,741	98.26
820	62	0.11	54,803	98.37
821	53	0.10	54,856	98.47
822	48	0.09	54,904	98.55
823	38	0.07	54,942	98.62
824	43	0.08	54,985	98.70
825	45	0.08	55,030	98.78
826	56	0.10	55,086	98.88
827	49	0.09	55,135	98.97
828	30	0.05	55,165	99.02
829	37	0.07	55,202	99.09
830	27	0.05	55,229	99.14
831	40	0.07	55,269	99.21
832	24	0.04	55,293	99.25
833	39	0.07	55,332	99.32
834	25	0.04	55,357	99.37
835	33	0.06	55,390	99.43
836	19	0.03	55,409	99.46
837	22	0.04	55,431	99.50
838	24	0.04	55,455	99.54
839	24	0.04	55,479	99.59
840	19	0.03	55,498	99.62
841	16	0.03	55,514	99.65
842	14	0.03	55,528	99.67
843	11	0.02	55,539	99.69
844	17	0.03	55,556	99.72
845	12	0.02	55,568	99.75
846	12	0.02	55,580	99.77
847	7	0.01	55,587	99.78
848	12	0.02	55,599	99.80
849	7	0.01	55,606	99.81
850	104	0.19	55,710	100.00

Table E.9. Scale Score Distribution—ELA Grade 5

SS	Freq.	%	Cum. Freq.	Cum. %
650	115	0.21	115	0.21
651	7	0.01	122	0.22
652	2	0.00	124	0.22
653	9	0.02	133	0.24
654	11	0.02	144	0.26
655	4	0.01	148	0.26
656	6	0.01	154	0.28
657	22	0.04	176	0.31
658	20	0.04	196	0.35
659	15	0.03	211	0.38
660	6	0.01	217	0.39
661	24	0.04	241	0.43
662	11	0.02	252	0.45
663	14	0.03	266	0.48
664	16	0.03	282	0.50
665	12	0.02	294	0.53
666	30	0.05	324	0.58
667	23	0.04	347	0.62
668	31	0.06	378	0.68
669	33	0.06	411	0.73
670	38	0.07	449	0.80
671	41	0.07	490	0.88
672	46	0.08	536	0.96
673	51	0.09	587	1.05
674	53	0.09	640	1.14
675	76	0.14	716	1.28
676	41	0.07	757	1.35
677	73	0.13	830	1.48
678	53	0.09	883	1.58
679	70	0.13	953	1.70
680	95	0.17	1,048	1.87
681	85	0.15	1,133	2.02
682	91	0.16	1,224	2.19
683	115	0.21	1,339	2.39
684	136	0.24	1,475	2.64
685	126	0.23	1,601	2.86
686	145	0.26	1,746	3.12
687	171	0.31	1,917	3.43
688	148	0.26	2,065	3.69
689	177	0.32	2,242	4.01
690	145	0.26	2,387	4.27
691	181	0.32	2,568	4.59
692	195	0.35	2,763	4.94
693	223	0.40	2,986	5.34
694	249	0.45	3,235	5.78
695	232	0.41	3,467	6.20
696	229	0.41	3,696	6.61
697	254	0.45	3,950	7.06
698	284	0.51	4,234	7.57
699	331	0.59	4,565	8.16
700	309	0.55	4,874	8.71
701	316	0.56	5,190	9.28
702	296	0.53	5,486	9.80
703	351	0.63	5,837	10.43
704	333	0.60	6,170	11.03
705	338	0.60	6,508	11.63
706	351	0.63	6,859	12.26
707	356	0.64	7,215	12.89
708	376	0.67	7,591	13.57
709	385	0.69	7,976	14.25
710	382	0.68	8,358	14.94
711	369	0.66	8,727	15.60
712	400	0.71	9,127	16.31
713	420	0.75	9,547	17.06
714	406	0.73	9,953	17.79
715	406	0.73	10,359	18.51
716	410	0.73	10,769	19.25
717	413	0.74	11,182	19.98
718	481	0.86	11,663	20.84
719	473	0.85	12,136	21.69
720	453	0.81	12,589	22.50
721	465	0.83	13,054	23.33
722	525	0.94	13,579	24.27
723	492	0.88	14,071	25.15
724	483	0.86	14,554	26.01
725	473	0.85	15,027	26.86
726	554	0.99	15,581	27.85
727	525	0.94	16,106	28.78
728	520	0.93	16,626	29.71
729	549	0.98	17,175	30.70
730	576	1.03	17,751	31.72
731	584	1.04	18,335	32.77
732	567	1.01	18,902	33.78
733	590	1.05	19,492	34.84
734	572	1.02	20,064	35.86
735	613	1.10	20,677	36.95
736	601	1.07	21,278	38.03
737	625	1.12	21,903	39.15
738	616	1.10	22,519	40.25
739	623	1.11	23,142	41.36
740	612	1.09	23,754	42.45
741	672	1.20	24,426	43.65
742	673	1.20	25,099	44.86
743	629	1.12	25,728	45.98
744	639	1.14	26,367	47.12
745	622	1.11	26,989	48.24
746	635	1.13	27,624	49.37
747	634	1.13	28,258	50.50
748	603	1.08	28,861	51.58
749	633	1.13	29,494	52.71

SS	Freq.	%	Cum. Freq.	Cum. %
750	605	1.08	30,099	53.79
751	645	1.15	30,744	54.95
752	633	1.13	31,377	56.08
753	627	1.12	32,004	57.20
754	636	1.14	32,640	58.33
755	634	1.13	33,274	59.47
756	645	1.15	33,919	60.62
757	612	1.09	34,531	61.71
758	617	1.10	35,148	62.82
759	631	1.13	35,779	63.94
760	597	1.07	36,376	65.01
761	596	1.07	36,972	66.08
762	612	1.09	37,584	67.17
763	560	1.00	38,144	68.17
764	558	1.00	38,702	69.17
765	585	1.05	39,287	70.21
766	602	1.08	39,889	71.29
767	552	0.99	40,441	72.28
768	538	0.96	40,979	73.24
769	531	0.95	41,510	74.19
770	530	0.95	42,040	75.13
771	540	0.97	42,580	76.10
772	474	0.85	43,054	76.95
773	477	0.85	43,531	77.80
774	496	0.89	44,027	78.69
775	467	0.83	44,494	79.52
776	468	0.84	44,962	80.36
777	464	0.83	45,426	81.19
778	435	0.78	45,861	81.96
779	456	0.81	46,317	82.78
780	422	0.75	46,739	83.53
781	411	0.73	47,150	84.27
782	369	0.66	47,519	84.93
783	395	0.71	47,914	85.63
784	387	0.69	48,301	86.32
785	363	0.65	48,664	86.97
786	352	0.63	49,016	87.60
787	352	0.63	49,368	88.23
788	333	0.60	49,701	88.83
789	315	0.56	50,016	89.39
790	324	0.58	50,340	89.97
791	305	0.55	50,645	90.51
792	277	0.50	50,922	91.01
793	280	0.50	51,202	91.51
794	231	0.41	51,433	91.92
795	251	0.45	51,684	92.37
796	239	0.43	51,923	92.80
797	238	0.43	52,161	93.22
798	181	0.32	52,342	93.55
799	197	0.35	52,539	93.90
800	215	0.38	52,754	94.28
801	203	0.36	52,957	94.65
802	152	0.27	53,109	94.92
803	178	0.32	53,287	95.24
804	185	0.33	53,472	95.57
805	162	0.29	53,634	95.86
806	155	0.28	53,789	96.13
807	129	0.23	53,918	96.36
808	136	0.24	54,054	96.61
809	109	0.19	54,163	96.80
810	106	0.19	54,269	96.99
811	116	0.21	54,385	97.20
812	98	0.18	54,483	97.37
813	87	0.16	54,570	97.53
814	90	0.16	54,660	97.69
815	83	0.15	54,743	97.84
816	89	0.16	54,832	98.00
817	68	0.12	54,900	98.12
818	65	0.12	54,965	98.23
819	80	0.14	55,045	98.38
820	57	0.10	55,102	98.48
821	64	0.11	55,166	98.59
822	52	0.09	55,218	98.69
823	50	0.09	55,268	98.78
824	43	0.08	55,311	98.85
825	40	0.07	55,351	98.92
826	33	0.06	55,384	98.98
827	41	0.07	55,425	99.06
828	38	0.07	55,463	99.12
829	45	0.08	55,508	99.20
830	33	0.06	55,541	99.26
831	24	0.04	55,565	99.31
832	38	0.07	55,603	99.37
833	19	0.03	55,622	99.41
834	23	0.04	55,645	99.45
835	24	0.04	55,669	99.49
836	21	0.04	55,690	99.53
837	22	0.04	55,712	99.57
838	23	0.04	55,735	99.61
839	16	0.03	55,751	99.64
840	13	0.02	55,764	99.66
841	14	0.03	55,778	99.69
842	11	0.02	55,789	99.71
843	9	0.02	55,798	99.72
844	10	0.02	55,808	99.74
845	11	0.02	55,819	99.76
846	5	0.01	55,824	99.77
847	12	0.02	55,836	99.79
848	11	0.02	55,847	99.81
849	8	0.01	55,855	99.82
850	98	0.18	55,953	100.00

Table E.10. Scale Score Distribution—ELA Grade 6

SS	Freq.	%	Cum. Freq.	Cum. %
650	90	0.16	90	0.16
651	14	0.03	104	0.19
652	13	0.02	117	0.21
653	9	0.02	126	0.23
654	4	0.01	130	0.24
655	8	0.01	138	0.25
656	9	0.02	147	0.27
657	11	0.02	158	0.29
658	7	0.01	165	0.30
659	11	0.02	176	0.32
660	22	0.04	198	0.36
661	26	0.05	224	0.41
662	23	0.04	247	0.45
663	21	0.04	268	0.49
664	23	0.04	291	0.53
665	20	0.04	311	0.57
666	30	0.05	341	0.62
667	38	0.07	379	0.69
668	38	0.07	417	0.76
669	53	0.10	470	0.86
670	53	0.10	523	0.96
671	52	0.10	575	1.05
672	42	0.08	617	1.13
673	60	0.11	677	1.24
674	61	0.11	738	1.35
675	82	0.15	820	1.50
676	88	0.16	908	1.66
677	86	0.16	994	1.82
678	87	0.16	1,081	1.98
679	100	0.18	1,181	2.16
680	125	0.23	1,306	2.39
681	140	0.26	1,446	2.65
682	132	0.24	1,578	2.89
683	129	0.24	1,707	3.13
684	140	0.26	1,847	3.39
685	161	0.30	2,008	3.68
686	155	0.28	2,163	3.96
687	184	0.34	2,347	4.30
688	212	0.39	2,559	4.69
689	215	0.39	2,774	5.08
690	213	0.39	2,987	5.47
691	230	0.42	3,217	5.90
692	275	0.50	3,492	6.40
693	261	0.48	3,753	6.88
694	277	0.51	4,030	7.39
695	296	0.54	4,326	7.93
696	287	0.53	4,613	8.45
697	286	0.52	4,899	8.98
698	338	0.62	5,237	9.60
699	304	0.56	5,541	10.16
700	348	0.64	5,889	10.79
701	373	0.68	6,262	11.48
702	356	0.65	6,618	12.13
703	383	0.70	7,001	12.83
704	369	0.68	7,370	13.51
705	360	0.66	7,730	14.17
706	398	0.73	8,128	14.90
707	434	0.80	8,562	15.69
708	392	0.72	8,954	16.41
709	418	0.77	9,372	17.18
710	451	0.83	9,823	18.00
711	398	0.73	10,221	18.73
712	450	0.82	10,671	19.56
713	431	0.79	11,102	20.35
714	460	0.84	11,562	21.19
715	446	0.82	12,008	22.01
716	484	0.89	12,492	22.90
717	440	0.81	12,932	23.70
718	471	0.86	13,403	24.56
719	438	0.80	13,841	25.37
720	455	0.83	14,296	26.20
721	470	0.86	14,766	27.06
722	541	0.99	15,307	28.05
723	520	0.95	15,827	29.01
724	516	0.95	16,343	29.95
725	523	0.96	16,866	30.91
726	527	0.97	17,393	31.88
727	516	0.95	17,909	32.82
728	570	1.04	18,479	33.87
729	536	0.98	19,015	34.85
730	539	0.99	19,554	35.84
731	547	1.00	20,101	36.84
732	523	0.96	20,624	37.80
733	596	1.09	21,220	38.89
734	544	1.00	21,764	39.89
735	538	0.99	22,302	40.87
736	559	1.02	22,861	41.90
737	545	1.00	23,406	42.90
738	555	1.02	23,961	43.92
739	633	1.16	24,594	45.08
740	601	1.10	25,195	46.18
741	576	1.06	25,771	47.23
742	570	1.04	26,341	48.28
743	605	1.11	26,946	49.39
744	603	1.11	27,549	50.49
745	598	1.10	28,147	51.59
746	569	1.04	28,716	52.63
747	594	1.09	29,310	53.72
748	606	1.11	29,916	54.83
749	641	1.17	30,557	56.00

SS	Freq.	%	Cum. Freq.	Cum. %
750	653	1.20	31,210	57.20
751	599	1.10	31,809	58.30
752	647	1.19	32,456	59.48
753	585	1.07	33,041	60.56
754	647	1.19	33,688	61.74
755	636	1.17	34,324	62.91
756	596	1.09	34,920	64.00
757	629	1.15	35,549	65.15
758	612	1.12	36,161	66.28
759	565	1.04	36,726	67.31
760	640	1.17	37,366	68.48
761	578	1.06	37,944	69.54
762	596	1.09	38,540	70.64
763	572	1.05	39,112	71.68
764	564	1.03	39,676	72.72
765	542	0.99	40,218	73.71
766	565	1.04	40,783	74.75
767	545	1.00	41,328	75.75
768	520	0.95	41,848	76.70
769	542	0.99	42,390	77.69
770	528	0.97	42,918	78.66
771	515	0.94	43,433	79.60
772	489	0.90	43,922	80.50
773	490	0.90	44,412	81.40
774	441	0.81	44,853	82.21
775	426	0.78	45,279	82.99
776	440	0.81	45,719	83.79
777	462	0.85	46,181	84.64
778	445	0.82	46,626	85.46
779	405	0.74	47,031	86.20
780	414	0.76	47,445	86.96
781	371	0.68	47,816	87.64
782	357	0.65	48,173	88.29
783	375	0.69	48,548	88.98
784	356	0.65	48,904	89.63
785	294	0.54	49,198	90.17
786	281	0.52	49,479	90.68
787	289	0.53	49,768	91.21
788	256	0.47	50,024	91.68
789	268	0.49	50,292	92.17
790	242	0.44	50,534	92.62
791	270	0.49	50,804	93.11
792	233	0.43	51,037	93.54
793	214	0.39	51,251	93.93
794	216	0.40	51,467	94.33
795	193	0.35	51,660	94.68
796	208	0.38	51,868	95.06
797	183	0.34	52,051	95.40
798	168	0.31	52,219	95.71
799	168	0.31	52,387	96.01
800	169	0.31	52,556	96.32
801	151	0.28	52,707	96.60
802	117	0.21	52,824	96.81
803	109	0.20	52,933	97.01
804	102	0.19	53,035	97.20
805	130	0.24	53,165	97.44
806	111	0.20	53,276	97.64
807	84	0.15	53,360	97.80
808	85	0.16	53,445	97.95
809	78	0.14	53,523	98.10
810	88	0.16	53,611	98.26
811	77	0.14	53,688	98.40
812	53	0.10	53,741	98.50
813	61	0.11	53,802	98.61
814	63	0.12	53,865	98.72
815	52	0.10	53,917	98.82
816	57	0.10	53,974	98.92
817	35	0.06	54,009	98.99
818	50	0.09	54,059	99.08
819	40	0.07	54,099	99.15
820	44	0.08	54,143	99.23
821	26	0.05	54,169	99.28
822	29	0.05	54,198	99.33
823	23	0.04	54,221	99.38
824	26	0.05	54,247	99.42
825	23	0.04	54,270	99.46
826	23	0.04	54,293	99.51
827	20	0.04	54,313	99.54
828	21	0.04	54,334	99.58
829	16	0.03	54,350	99.61
830	21	0.04	54,371	99.65
831	13	0.02	54,384	99.67
832	10	0.02	54,394	99.69
833	13	0.02	54,407	99.72
834	12	0.02	54,419	99.74
835	16	0.03	54,435	99.77
836	10	0.02	54,445	99.79
837	13	0.02	54,458	99.81
838	9	0.02	54,467	99.83
839	10	0.02	54,477	99.84
840	11	0.02	54,488	99.86
841	3	0.01	54,491	99.87
842	13	0.02	54,504	99.89
843	8	0.01	54,512	99.91
844	2	0.00	54,514	99.91
845	4	0.01	54,518	99.92
847	3	0.01	54,521	99.92
848	3	0.01	54,524	99.93
849	4	0.01	54,528	99.94
850	34	0.06	54,562	100.00

Table E.11. Scale Score Distribution—ELA Grade 7

SS	Freq.	%	Cum. Freq.	Cum. %
650	153	0.29	153	0.29
651	13	0.02	166	0.31
652	9	0.02	175	0.33
653	16	0.03	191	0.36
654	14	0.03	205	0.38
655	21	0.04	226	0.42
656	18	0.03	244	0.46
657	24	0.04	268	0.50
658	23	0.04	291	0.55
659	25	0.05	316	0.59
660	28	0.05	344	0.64
661	37	0.07	381	0.71
662	33	0.06	414	0.78
663	38	0.07	452	0.85
664	48	0.09	500	0.94
665	54	0.10	554	1.04
666	59	0.11	613	1.15
667	66	0.12	679	1.27
668	57	0.11	736	1.38
669	66	0.12	802	1.50
670	78	0.15	880	1.65
671	84	0.16	964	1.81
672	72	0.13	1,036	1.94
673	96	0.18	1,132	2.12
674	104	0.19	1,236	2.32
675	102	0.19	1,338	2.51
676	102	0.19	1,440	2.70
677	115	0.22	1,555	2.92
678	151	0.28	1,706	3.20
679	125	0.23	1,831	3.43
680	156	0.29	1,987	3.73
681	142	0.27	2,129	3.99
682	167	0.31	2,296	4.30
683	171	0.32	2,467	4.63
684	197	0.37	2,664	4.99
685	231	0.43	2,895	5.43
686	226	0.42	3,121	5.85
687	214	0.40	3,335	6.25
688	217	0.41	3,552	6.66
689	244	0.46	3,796	7.12
690	256	0.48	4,052	7.60
691	236	0.44	4,288	8.04
692	273	0.51	4,561	8.55
693	315	0.59	4,876	9.14
694	308	0.58	5,184	9.72
695	327	0.61	5,511	10.33
696	318	0.60	5,829	10.93
697	312	0.58	6,141	11.51
698	358	0.67	6,499	12.19
699	333	0.62	6,832	12.81
700	337	0.63	7,169	13.44
701	311	0.58	7,480	14.02
702	353	0.66	7,833	14.69
703	349	0.65	8,182	15.34
704	355	0.67	8,537	16.01
705	352	0.66	8,889	16.67
706	395	0.74	9,284	17.41
707	395	0.74	9,679	18.15
708	368	0.69	10,047	18.84
709	332	0.62	10,379	19.46
710	388	0.73	10,767	20.19
711	430	0.81	11,197	20.99
712	367	0.69	11,564	21.68
713	384	0.72	11,948	22.40
714	396	0.74	12,344	23.14
715	391	0.73	12,735	23.88
716	442	0.83	13,177	24.71
717	401	0.75	13,578	25.46
718	400	0.75	13,978	26.21
719	398	0.75	14,376	26.95
720	431	0.81	14,807	27.76
721	421	0.79	15,228	28.55
722	438	0.82	15,666	29.37
723	408	0.76	16,074	30.14
724	416	0.78	16,490	30.92
725	420	0.79	16,910	31.71
726	430	0.81	17,340	32.51
727	475	0.89	17,815	33.40
728	460	0.86	18,275	34.27
729	455	0.85	18,730	35.12
730	492	0.92	19,222	36.04
731	500	0.94	19,722	36.98
732	491	0.92	20,213	37.90
733	452	0.85	20,665	38.75
734	468	0.88	21,133	39.62
735	466	0.87	21,599	40.50
736	478	0.90	22,077	41.39
737	464	0.87	22,541	42.26
738	491	0.92	23,032	43.18
739	510	0.96	23,542	44.14
740	492	0.92	24,034	45.06
741	514	0.96	24,548	46.03
742	488	0.91	25,036	46.94
743	505	0.95	25,541	47.89
744	489	0.92	26,030	48.81
745	531	1.00	26,561	49.80
746	497	0.93	27,058	50.73
747	528	0.99	27,586	51.72
748	511	0.96	28,097	52.68
749	529	0.99	28,626	53.67

SS	Freq.	%	Cum. Freq.	Cum. %
750	526	0.99	29,152	54.66
751	497	0.93	29,649	55.59
752	502	0.94	30,151	56.53
753	460	0.86	30,611	57.39
754	528	0.99	31,139	58.38
755	469	0.88	31,608	59.26
756	506	0.95	32,114	60.21
757	472	0.88	32,586	61.10
758	462	0.87	33,048	61.96
759	461	0.86	33,509	62.83
760	498	0.93	34,007	63.76
761	475	0.89	34,482	64.65
762	457	0.86	34,939	65.51
763	504	0.94	35,443	66.45
764	429	0.80	35,872	67.26
765	497	0.93	36,369	68.19
766	473	0.89	36,842	69.08
767	467	0.88	37,309	69.95
768	466	0.87	37,775	70.83
769	449	0.84	38,224	71.67
770	461	0.86	38,685	72.53
771	436	0.82	39,121	73.35
772	432	0.81	39,553	74.16
773	410	0.77	39,963	74.93
774	413	0.77	40,376	75.70
775	404	0.76	40,780	76.46
776	385	0.72	41,165	77.18
777	412	0.77	41,577	77.96
778	396	0.74	41,973	78.70
779	386	0.72	42,359	79.42
780	396	0.74	42,755	80.16
781	371	0.70	43,126	80.86
782	338	0.63	43,464	81.49
783	396	0.74	43,860	82.24
784	368	0.69	44,228	82.93
785	366	0.69	44,594	83.61
786	356	0.67	44,950	84.28
787	330	0.62	45,280	84.90
788	315	0.59	45,595	85.49
789	320	0.60	45,915	86.09
790	321	0.60	46,236	86.69
791	289	0.54	46,525	87.23
792	296	0.55	46,821	87.79
793	279	0.52	47,100	88.31
794	263	0.49	47,363	88.80
795	283	0.53	47,646	89.34
796	239	0.45	47,885	89.78
797	246	0.46	48,131	90.24
798	227	0.43	48,358	90.67
799	251	0.47	48,609	91.14
800	246	0.46	48,855	91.60
801	217	0.41	49,072	92.01
802	218	0.41	49,290	92.42
803	223	0.42	49,513	92.84
804	199	0.37	49,712	93.21
805	178	0.33	49,890	93.54
806	183	0.34	50,073	93.89
807	164	0.31	50,237	94.19
808	159	0.30	50,396	94.49
809	155	0.29	50,551	94.78
810	141	0.26	50,692	95.05
811	155	0.29	50,847	95.34
812	127	0.24	50,974	95.58
813	133	0.25	51,107	95.82
814	132	0.25	51,239	96.07
815	121	0.23	51,360	96.30
816	113	0.21	51,473	96.51
817	97	0.18	51,570	96.69
818	102	0.19	51,672	96.88
819	97	0.18	51,769	97.07
820	76	0.14	51,845	97.21
821	89	0.17	51,934	97.38
822	84	0.16	52,018	97.53
823	86	0.16	52,104	97.69
824	77	0.14	52,181	97.84
825	86	0.16	52,267	98.00
826	53	0.10	52,320	98.10
827	69	0.13	52,389	98.23
828	51	0.10	52,440	98.32
829	57	0.11	52,497	98.43
830	51	0.10	52,548	98.53
831	40	0.07	52,588	98.60
832	46	0.09	52,634	98.69
833	42	0.08	52,676	98.77
834	39	0.07	52,715	98.84
835	38	0.07	52,753	98.91
836	41	0.08	52,794	98.99
837	34	0.06	52,828	99.05
838	42	0.08	52,870	99.13
839	29	0.05	52,899	99.18
840	26	0.05	52,925	99.23
841	31	0.06	52,956	99.29
842	24	0.04	52,980	99.34
843	21	0.04	53,001	99.38
844	22	0.04	53,023	99.42
845	22	0.04	53,045	99.46
846	17	0.03	53,062	99.49
847	20	0.04	53,082	99.53
848	18	0.03	53,100	99.56
849	16	0.03	53,116	99.59
850	218	0.41	53,334	100.00

Table E.12. Scale Score Distribution—ELA Grade 8

SS	Freq.	%	Cum. Freq.	Cum. %
650	620	1.23	620	1.23
651	41	0.08	661	1.31
652	42	0.08	703	1.40
653	58	0.12	761	1.51
654	67	0.13	828	1.64
655	58	0.12	886	1.76
656	57	0.11	943	1.87
657	73	0.15	1,016	2.02
658	82	0.16	1,098	2.18
659	83	0.16	1,181	2.35
660	85	0.17	1,266	2.51
661	83	0.16	1,349	2.68
662	88	0.17	1,437	2.85
663	126	0.25	1,563	3.10
664	131	0.26	1,694	3.37
665	123	0.24	1,817	3.61
666	108	0.21	1,925	3.82
667	122	0.24	2,047	4.07
668	134	0.27	2,181	4.33
669	141	0.28	2,322	4.61
670	152	0.30	2,474	4.91
671	152	0.30	2,626	5.22
672	170	0.34	2,796	5.55
673	158	0.31	2,954	5.87
674	177	0.35	3,131	6.22
675	179	0.36	3,310	6.58
676	197	0.39	3,507	6.97
677	191	0.38	3,698	7.35
678	189	0.38	3,887	7.72
679	232	0.46	4,119	8.18
680	227	0.45	4,346	8.63
681	228	0.45	4,574	9.09
682	211	0.42	4,785	9.51
683	223	0.44	5,008	9.95
684	239	0.47	5,247	10.42
685	231	0.46	5,478	10.88
686	233	0.46	5,711	11.34
687	233	0.46	5,944	11.81
688	289	0.57	6,233	12.38
689	267	0.53	6,500	12.91
690	232	0.46	6,732	13.37
691	256	0.51	6,988	13.88
692	280	0.56	7,268	14.44
693	261	0.52	7,529	14.96
694	243	0.48	7,772	15.44
695	255	0.51	8,027	15.95
696	271	0.54	8,298	16.48
697	259	0.51	8,557	17.00
698	259	0.51	8,816	17.51
699	279	0.55	9,095	18.07
700	290	0.58	9,385	18.64
701	292	0.58	9,677	19.22
702	293	0.58	9,970	19.80
703	288	0.57	10,258	20.38
704	321	0.64	10,579	21.01
705	295	0.59	10,874	21.60
706	352	0.70	11,226	22.30
707	298	0.59	11,524	22.89
708	319	0.63	11,843	23.53
709	331	0.66	12,174	24.18
710	320	0.64	12,494	24.82
711	351	0.70	12,845	25.52
712	329	0.65	13,174	26.17
713	322	0.64	13,496	26.81
714	360	0.72	13,856	27.52
715	350	0.70	14,206	28.22
716	371	0.74	14,577	28.96
717	351	0.70	14,928	29.65
718	355	0.71	15,283	30.36
719	361	0.72	15,644	31.08
720	407	0.81	16,051	31.88
721	399	0.79	16,450	32.68
722	368	0.73	16,818	33.41
723	410	0.81	17,228	34.22
724	374	0.74	17,602	34.97
725	423	0.84	18,025	35.81
726	413	0.82	18,438	36.63
727	444	0.88	18,882	37.51
728	395	0.78	19,277	38.29
729	427	0.85	19,704	39.14
730	440	0.87	20,144	40.02
731	438	0.87	20,582	40.89
732	416	0.83	20,998	41.71
733	438	0.87	21,436	42.58
734	413	0.82	21,849	43.40
735	424	0.84	22,273	44.24
736	441	0.88	22,714	45.12
737	450	0.89	23,164	46.01
738	473	0.94	23,637	46.95
739	494	0.98	24,131	47.94
740	485	0.96	24,616	48.90
741	463	0.92	25,079	49.82
742	468	0.93	25,547	50.75
743	447	0.89	25,994	51.64
744	464	0.92	26,458	52.56
745	472	0.94	26,930	53.50
746	461	0.92	27,391	54.41
747	523	1.04	27,914	55.45
748	452	0.90	28,366	56.35
749	427	0.85	28,793	57.20

SS	Freq.	%	Cum. Freq.	Cum. %
750	492	0.98	29,285	58.17
751	475	0.94	29,760	59.12
752	471	0.94	30,231	60.05
753	479	0.95	30,710	61.00
754	440	0.87	31,150	61.88
755	459	0.91	31,609	62.79
756	443	0.88	32,052	63.67
757	499	0.99	32,551	64.66
758	505	1.00	33,056	65.66
759	466	0.93	33,522	66.59
760	447	0.89	33,969	67.48
761	455	0.90	34,424	68.38
762	415	0.82	34,839	69.21
763	439	0.87	35,278	70.08
764	440	0.87	35,718	70.95
765	418	0.83	36,136	71.78
766	402	0.80	36,538	72.58
767	354	0.70	36,892	73.28
768	413	0.82	37,305	74.10
769	443	0.88	37,748	74.98
770	376	0.75	38,124	75.73
771	420	0.83	38,544	76.57
772	430	0.85	38,974	77.42
773	408	0.81	39,382	78.23
774	423	0.84	39,805	79.07
775	394	0.78	40,199	79.85
776	347	0.69	40,546	80.54
777	353	0.70	40,899	81.24
778	320	0.64	41,219	81.88
779	355	0.71	41,574	82.58
780	361	0.72	41,935	83.30
781	296	0.59	42,231	83.89
782	306	0.61	42,537	84.50
783	283	0.56	42,820	85.06
784	324	0.64	43,144	85.70
785	274	0.54	43,418	86.25
786	308	0.61	43,726	86.86
787	259	0.51	43,985	87.37
788	256	0.51	44,241	87.88
789	251	0.50	44,492	88.38
790	264	0.52	44,756	88.91
791	238	0.47	44,994	89.38
792	249	0.49	45,243	89.87
793	195	0.39	45,438	90.26
794	230	0.46	45,668	90.72
795	208	0.41	45,876	91.13
796	219	0.44	46,095	91.57
797	227	0.45	46,322	92.02
798	218	0.43	46,540	92.45
799	197	0.39	46,737	92.84
800	164	0.33	46,901	93.17
801	155	0.31	47,056	93.47
802	172	0.34	47,228	93.82
803	157	0.31	47,385	94.13
804	147	0.29	47,532	94.42
805	136	0.27	47,668	94.69
806	136	0.27	47,804	94.96
807	138	0.27	47,942	95.23
808	130	0.26	48,072	95.49
809	137	0.27	48,209	95.76
810	89	0.18	48,298	95.94
811	95	0.19	48,393	96.13
812	95	0.19	48,488	96.32
813	92	0.18	48,580	96.50
814	106	0.21	48,686	96.71
815	101	0.20	48,787	96.91
816	78	0.15	48,865	97.07
817	82	0.16	48,947	97.23
818	93	0.18	49,040	97.42
819	70	0.14	49,110	97.55
820	76	0.15	49,186	97.71
821	60	0.12	49,246	97.82
822	66	0.13	49,312	97.96
823	67	0.13	49,379	98.09
824	54	0.11	49,433	98.20
825	63	0.13	49,496	98.32
826	68	0.14	49,564	98.46
827	52	0.10	49,616	98.56
828	42	0.08	49,658	98.64
829	45	0.09	49,703	98.73
830	50	0.10	49,753	98.83
831	42	0.08	49,795	98.92
832	50	0.10	49,845	99.01
833	24	0.05	49,869	99.06
834	36	0.07	49,905	99.13
835	23	0.05	49,928	99.18
836	25	0.05	49,953	99.23
837	31	0.06	49,984	99.29
838	29	0.06	50,013	99.35
839	24	0.05	50,037	99.40
840	22	0.04	50,059	99.44
841	16	0.03	50,075	99.47
842	22	0.04	50,097	99.52
843	12	0.02	50,109	99.54
844	22	0.04	50,131	99.58
845	15	0.03	50,146	99.61
846	17	0.03	50,163	99.65
847	14	0.03	50,177	99.67
848	15	0.03	50,192	99.70
849	14	0.03	50,206	99.73
850	135	0.27	50,341	100.00

Table E.13. Scale Score Distribution—CSLA Grade 3

SS	Freq.	%	Cum. Freq.	Cum. %
650	24	1.54	24	1.54
667	26	1.67	50	3.21
678	41	2.63	91	5.83
686	56	3.59	147	9.42
692	58	3.72	205	13.14
697	61	3.91	266	17.05
700	72	4.62	338	21.67
705	61	3.91	399	25.58
708	74	4.74	473	30.32
711	62	3.97	535	34.29
714	52	3.33	587	37.63
717	57	3.65	644	41.28
719	41	2.63	685	43.91
721	51	3.27	736	47.18
723	50	3.21	786	50.38
725	47	3.01	833	53.40
727	48	3.08	881	56.47
729	38	2.44	919	58.91
730	30	1.92	949	60.83
732	22	1.41	971	62.24
734	43	2.76	1,014	65.00
735	33	2.12	1,047	67.12
737	31	1.99	1,078	69.10
738	29	1.86	1,107	70.96
740	28	1.79	1,135	72.76
741	24	1.54	1,159	74.29
743	29	1.86	1,188	76.15
744	19	1.22	1,207	77.37
746	24	1.54	1,231	78.91
747	31	1.99	1,262	80.90
748	29	1.86	1,291	82.76
750	22	1.41	1,313	84.17
751	20	1.28	1,333	85.45
753	32	2.05	1,365	87.50
754	16	1.03	1,381	88.53
755	21	1.35	1,402	89.87
757	21	1.35	1,423	91.22
758	22	1.41	1,445	92.63
760	9	0.58	1,454	93.21
761	8	0.51	1,462	93.72
763	11	0.71	1,473	94.42
764	11	0.71	1,484	95.13
766	6	0.38	1,490	95.51
767	15	0.96	1,505	96.47
769	7	0.45	1,512	96.92
771	3	0.19	1,515	97.12
772	6	0.38	1,521	97.50
774	10	0.64	1,531	98.14
776	4	0.26	1,535	98.40
778	2	0.13	1,537	98.53
779	2	0.13	1,539	98.65
782	6	0.38	1,545	99.04
785	5	0.32	1,550	99.36
787	1	0.06	1,551	99.42
790	2	0.13	1,553	99.55
793	2	0.13	1,555	99.68
796	1	0.06	1,556	99.74
800	1	0.06	1,557	99.81
804	1	0.06	1,558	99.87
815	1	0.06	1,559	99.94
823	1	0.06	1,560	100.00

Table E.14. Scale Score Distribution—CSLA Grade 4

SS	Freq.	%	Cum. Freq.	Cum. %
650	7	0.60	7	0.60
654	8	0.69	15	1.29
665	26	2.24	41	3.53
672	28	2.41	69	5.94
678	41	3.53	110	9.47
683	37	3.18	147	12.65
687	35	3.01	182	15.66
691	46	3.96	228	19.62
694	31	2.67	259	22.29
697	34	2.93	293	25.22
700	24	2.07	317	27.28
703	38	3.27	355	30.55
705	26	2.24	381	32.79
707	25	2.15	406	34.94
709	24	2.07	430	37.01
711	22	1.89	452	38.90
713	19	1.64	471	40.53
715	21	1.81	492	42.34
717	25	2.15	517	44.49
718	26	2.24	543	46.73
720	24	2.07	567	48.80
722	28	2.41	595	51.20
723	22	1.89	617	53.10
725	23	1.98	640	55.08
726	24	2.07	664	57.14
728	18	1.55	682	58.69
729	17	1.46	699	60.15
731	27	2.32	726	62.48
732	22	1.89	748	64.37
734	24	2.07	772	66.44
735	15	1.29	787	67.73
736	27	2.32	814	70.05
738	18	1.55	832	71.60
739	26	2.24	858	73.84
741	21	1.81	879	75.65
742	19	1.64	898	77.28
743	24	2.07	922	79.35
745	15	1.29	937	80.64
746	16	1.38	953	82.01
747	12	1.03	965	83.05
749	12	1.03	977	84.08
750	18	1.55	995	85.63
751	20	1.72	1,015	87.35
753	14	1.20	1,029	88.55
754	14	1.20	1,043	89.76
756	7	0.60	1,050	90.36
757	21	1.81	1,071	92.17
759	10	0.86	1,081	93.03
760	10	0.86	1,091	93.89
762	8	0.69	1,099	94.58
763	8	0.69	1,107	95.27
765	7	0.60	1,114	95.87
767	10	0.86	1,124	96.73
768	2	0.17	1,126	96.90
770	10	0.86	1,136	97.76
772	2	0.17	1,138	97.93
774	6	0.52	1,144	98.45
776	6	0.52	1,150	98.97
778	1	0.09	1,151	99.05
781	1	0.09	1,152	99.14
783	4	0.34	1,156	99.48
786	2	0.17	1,158	99.66
792	2	0.17	1,160	99.83
800	1	0.09	1,161	99.91
804	1	0.09	1,162	100.00

Table E.15. Scale Score Distribution—Science Grade 5

SS	Freq.	%	Cum. Freq.	Cum. %
650	976	1.74	976	1.74
651	64	0.11	1,040	1.85
652	43	0.08	1,083	1.93
653	54	0.10	1,137	2.03
654	44	0.08	1,181	2.11
655	52	0.09	1,233	2.20
656	43	0.08	1,276	2.28
657	44	0.08	1,320	2.35
658	50	0.09	1,370	2.44
659	49	0.09	1,419	2.53
660	72	0.13	1,491	2.66
661	45	0.08	1,536	2.74
662	79	0.14	1,615	2.88
663	74	0.13	1,689	3.01
664	76	0.14	1,765	3.15
665	79	0.14	1,844	3.29
666	71	0.13	1,915	3.41
667	83	0.15	1,998	3.56
668	113	0.20	2,111	3.76
669	117	0.21	2,228	3.97
670	107	0.19	2,335	4.16
671	113	0.20	2,448	4.36
672	101	0.18	2,549	4.54
673	93	0.17	2,642	4.71
674	129	0.23	2,771	4.94
675	117	0.21	2,888	5.15
676	145	0.26	3,033	5.41
677	138	0.25	3,171	5.65
678	134	0.24	3,305	5.89
679	161	0.29	3,466	6.18
680	157	0.28	3,623	6.46
681	142	0.25	3,765	6.71
682	136	0.24	3,901	6.96
683	179	0.32	4,080	7.27
684	187	0.33	4,267	7.61
685	177	0.32	4,444	7.92
686	196	0.35	4,640	8.27
687	209	0.37	4,849	8.65
688	218	0.39	5,067	9.03
689	199	0.35	5,266	9.39
690	212	0.38	5,478	9.77
691	248	0.44	5,726	10.21
692	229	0.41	5,955	10.62
693	263	0.47	6,218	11.09
694	243	0.43	6,461	11.52
695	249	0.44	6,710	11.96
696	259	0.46	6,969	12.43
697	269	0.48	7,238	12.91
698	311	0.55	7,549	13.46
699	330	0.59	7,879	14.05
700	286	0.51	8,165	14.56
701	316	0.56	8,481	15.12
702	325	0.58	8,806	15.70
703	353	0.63	9,159	16.33
704	326	0.58	9,485	16.91
705	339	0.60	9,824	17.52
706	367	0.65	10,191	18.17
707	370	0.66	10,561	18.83
708	375	0.67	10,936	19.50
709	397	0.71	11,333	20.21
710	399	0.71	11,732	20.92
711	427	0.76	12,159	21.68
712	435	0.78	12,594	22.46
713	457	0.81	13,051	23.27
714	424	0.76	13,475	24.03
715	440	0.78	13,915	24.81
716	503	0.90	14,418	25.71
717	503	0.90	14,921	26.60
718	474	0.85	15,395	27.45
719	467	0.83	15,862	28.28
720	510	0.91	16,372	29.19
721	533	0.95	16,905	30.14
722	566	1.01	17,471	31.15
723	545	0.97	18,016	32.12
724	555	0.99	18,571	33.11
725	563	1.00	19,134	34.12
726	545	0.97	19,679	35.09
727	614	1.09	20,293	36.18
728	583	1.04	20,876	37.22
729	627	1.12	21,503	38.34
730	610	1.09	22,113	39.43
731	611	1.09	22,724	40.52
732	563	1.00	23,287	41.52
733	673	1.20	23,960	42.72
734	632	1.13	24,592	43.85
735	695	1.24	25,287	45.09
736	623	1.11	25,910	46.20
737	655	1.17	26,565	47.37
738	694	1.24	27,259	48.60
739	686	1.22	27,945	49.83
740	626	1.12	28,571	50.94
741	652	1.16	29,223	52.10
742	661	1.18	29,884	53.28
743	700	1.25	30,584	54.53
744	701	1.25	31,285	55.78
745	704	1.26	31,989	57.04
746	686	1.22	32,675	58.26
747	748	1.33	33,423	59.59
748	694	1.24	34,117	60.83
749	699	1.25	34,816	62.08

SS	Freq.	%	Cum. Freq.	Cum. %
750	735	1.31	35,551	63.39
751	694	1.24	36,245	64.63
752	688	1.23	36,933	65.85
753	667	1.19	37,600	67.04
754	666	1.19	38,266	68.23
755	675	1.20	38,941	69.43
756	679	1.21	39,620	70.64
757	689	1.23	40,309	71.87
758	671	1.20	40,980	73.07
759	637	1.14	41,617	74.20
760	661	1.18	42,278	75.38
761	640	1.14	42,918	76.52
762	602	1.07	43,520	77.60
763	621	1.11	44,141	78.70
764	558	0.99	44,699	79.70
765	570	1.02	45,269	80.71
766	593	1.06	45,862	81.77
767	552	0.98	46,414	82.76
768	553	0.99	46,967	83.74
769	544	0.97	47,511	84.71
770	525	0.94	48,036	85.65
771	460	0.82	48,496	86.47
772	443	0.79	48,939	87.26
773	482	0.86	49,421	88.12
774	440	0.78	49,861	88.90
775	399	0.71	50,260	89.61
776	394	0.70	50,654	90.32
777	340	0.61	50,994	90.92
778	345	0.62	51,339	91.54
779	344	0.61	51,683	92.15
780	338	0.60	52,021	92.75
781	297	0.53	52,318	93.28
782	288	0.51	52,606	93.80
783	241	0.43	52,847	94.23
784	258	0.46	53,105	94.69
785	260	0.46	53,365	95.15
786	222	0.40	53,587	95.55
787	218	0.39	53,805	95.93
788	171	0.30	53,976	96.24
789	175	0.31	54,151	96.55
790	148	0.26	54,299	96.82
791	166	0.30	54,465	97.11
792	157	0.28	54,622	97.39
793	112	0.20	54,734	97.59
794	137	0.24	54,871	97.84
795	116	0.21	54,987	98.04
796	100	0.18	55,087	98.22
797	104	0.19	55,191	98.41
798	83	0.15	55,274	98.55
799	93	0.17	55,367	98.72
800	71	0.13	55,438	98.85
801	54	0.10	55,492	98.94
802	56	0.10	55,548	99.04
803	48	0.09	55,596	99.13
804	47	0.08	55,643	99.21
805	56	0.10	55,699	99.31
806	43	0.08	55,742	99.39
807	31	0.06	55,773	99.44
808	30	0.05	55,803	99.50
809	24	0.04	55,827	99.54
810	27	0.05	55,854	99.59
811	21	0.04	55,875	99.63
812	29	0.05	55,904	99.68
813	17	0.03	55,921	99.71
814	26	0.05	55,947	99.75
815	17	0.03	55,964	99.78
816	8	0.01	55,972	99.80
817	11	0.02	55,983	99.82
818	14	0.02	55,997	99.84
819	9	0.02	56,006	99.86
820	12	0.02	56,018	99.88
821	6	0.01	56,024	99.89
822	8	0.01	56,032	99.91
823	9	0.02	56,041	99.92
824	6	0.01	56,047	99.93
825	4	0.01	56,051	99.94
826	3	0.01	56,054	99.94
827	3	0.01	56,057	99.95
828	5	0.01	56,062	99.96
829	5	0.01	56,067	99.97
830	5	0.01	56,072	99.98
831	1	0.00	56,073	99.98
832	2	0.00	56,075	99.98
834	2	0.00	56,077	99.99
836	2	0.00	56,079	99.99
837	1	0.00	56,080	99.99
838	1	0.00	56,081	99.99
839	2	0.00	56,083	100.00
841	1	0.00	56,084	100.00
850	1	0.00	56,085	100.00

Table E.16. Scale Score Distribution—Science Grade 8

SS	Freq.	%	Cum. Freq.	Cum. %
650	2,137	4.28	2,137	4.28
651	29	0.06	2,166	4.34
652	40	0.08	2,206	4.42
653	38	0.08	2,244	4.50
654	25	0.05	2,269	4.55
655	28	0.06	2,297	4.60
656	29	0.06	2,326	4.66
657	57	0.11	2,383	4.77
658	60	0.12	2,443	4.89
659	56	0.11	2,499	5.01
660	62	0.12	2,561	5.13
661	67	0.13	2,628	5.27
662	65	0.13	2,693	5.40
663	65	0.13	2,758	5.53
664	71	0.14	2,829	5.67
665	72	0.14	2,901	5.81
666	92	0.18	2,993	6.00
667	78	0.16	3,071	6.15
668	87	0.17	3,158	6.33
669	105	0.21	3,263	6.54
670	119	0.24	3,382	6.78
671	116	0.23	3,498	7.01
672	120	0.24	3,618	7.25
673	115	0.23	3,733	7.48
674	96	0.19	3,829	7.67
675	116	0.23	3,945	7.90
676	112	0.22	4,057	8.13
677	129	0.26	4,186	8.39
678	137	0.27	4,323	8.66
679	156	0.31	4,479	8.97
680	133	0.27	4,612	9.24
681	141	0.28	4,753	9.52
682	165	0.33	4,918	9.85
683	153	0.31	5,071	10.16
684	163	0.33	5,234	10.49
685	185	0.37	5,419	10.86
686	210	0.42	5,629	11.28
687	222	0.44	5,851	11.72
688	214	0.43	6,065	12.15
689	207	0.41	6,272	12.57
690	234	0.47	6,506	13.03
691	228	0.46	6,734	13.49
692	228	0.46	6,962	13.95
693	237	0.47	7,199	14.42
694	235	0.47	7,434	14.89
695	256	0.51	7,690	15.41
696	252	0.50	7,942	15.91
697	256	0.51	8,198	16.42
698	285	0.57	8,483	17.00
699	290	0.58	8,773	17.58
700	306	0.61	9,079	18.19
701	326	0.65	9,405	18.84
702	288	0.58	9,693	19.42
703	300	0.60	9,993	20.02
704	325	0.65	10,318	20.67
705	348	0.70	10,666	21.37
706	338	0.68	11,004	22.05
707	378	0.76	11,382	22.80
708	366	0.73	11,748	23.54
709	388	0.78	12,136	24.31
710	397	0.80	12,533	25.11
711	379	0.76	12,912	25.87
712	383	0.77	13,295	26.64
713	430	0.86	13,725	27.50
714	447	0.90	14,172	28.39
715	424	0.85	14,596	29.24
716	435	0.87	15,031	30.11
717	455	0.91	15,486	31.03
718	455	0.91	15,941	31.94
719	466	0.93	16,407	32.87
720	454	0.91	16,861	33.78
721	467	0.94	17,328	34.72
722	510	1.02	17,838	35.74
723	472	0.95	18,310	36.68
724	522	1.05	18,832	37.73
725	515	1.03	19,347	38.76
726	560	1.12	19,907	39.88
727	544	1.09	20,451	40.97
728	533	1.07	20,984	42.04
729	566	1.13	21,550	43.17
730	597	1.20	22,147	44.37
731	518	1.04	22,665	45.41
732	626	1.25	23,291	46.66
733	586	1.17	23,877	47.84
734	625	1.25	24,502	49.09
735	605	1.21	25,107	50.30
736	576	1.15	25,683	51.45
737	590	1.18	26,273	52.64
738	602	1.21	26,875	53.84
739	613	1.23	27,488	55.07
740	657	1.32	28,145	56.39
741	618	1.24	28,763	57.63
742	601	1.20	29,364	58.83
743	622	1.25	29,986	60.08
744	632	1.27	30,618	61.34
745	637	1.28	31,255	62.62
746	640	1.28	31,895	63.90
747	636	1.27	32,531	65.17
748	647	1.30	33,178	66.47
749	641	1.28	33,819	67.75

SS	Freq.	%	Cum. Freq.	Cum. %
750	655	1.31	34,474	69.07
751	618	1.24	35,092	70.30
752	584	1.17	35,676	71.47
753	594	1.19	36,270	72.66
754	590	1.18	36,860	73.85
755	598	1.20	37,458	75.05
756	590	1.18	38,048	76.23
757	622	1.25	38,670	77.47
758	552	1.11	39,222	78.58
759	589	1.18	39,811	79.76
760	561	1.12	40,372	80.88
761	525	1.05	40,897	81.93
762	527	1.06	41,424	82.99
763	508	1.02	41,932	84.01
764	507	1.02	42,439	85.02
765	474	0.95	42,913	85.97
766	467	0.94	43,380	86.91
767	412	0.83	43,792	87.73
768	430	0.86	44,222	88.60
769	390	0.78	44,612	89.38
770	381	0.76	44,993	90.14
771	372	0.75	45,365	90.89
772	376	0.75	45,741	91.64
773	369	0.74	46,110	92.38
774	326	0.65	46,436	93.03
775	314	0.63	46,750	93.66
776	306	0.61	47,056	94.27
777	247	0.49	47,303	94.77
778	251	0.50	47,554	95.27
779	224	0.45	47,778	95.72
780	205	0.41	47,983	96.13
781	183	0.37	48,166	96.50
782	211	0.42	48,377	96.92
783	159	0.32	48,536	97.24
784	162	0.32	48,698	97.56
785	124	0.25	48,822	97.81

SS	Freq.	%	Cum. Freq.	Cum. %
786	134	0.27	48,956	98.08
787	107	0.21	49,063	98.30
788	81	0.16	49,144	98.46
789	78	0.16	49,222	98.61
790	80	0.16	49,302	98.77
791	91	0.18	49,393	98.96
792	66	0.13	49,459	99.09
793	59	0.12	49,518	99.21
794	63	0.13	49,581	99.33
795	42	0.08	49,623	99.42
796	28	0.06	49,651	99.47
797	44	0.09	49,695	99.56
798	32	0.06	49,727	99.63
799	22	0.04	49,749	99.67
800	24	0.05	49,773	99.72
801	26	0.05	49,799	99.77
802	20	0.04	49,819	99.81
803	13	0.03	49,832	99.84
804	11	0.02	49,843	99.86
805	12	0.02	49,855	99.88
806	12	0.02	49,867	99.91
807	5	0.01	49,872	99.92
808	11	0.02	49,883	99.94
809	2	0.00	49,885	99.94
810	4	0.01	49,889	99.95
811	4	0.01	49,893	99.96
812	1	0.00	49,894	99.96
813	3	0.01	49,897	99.97
814	4	0.01	49,901	99.97
815	2	0.00	49,903	99.98
816	4	0.01	49,907	99.99
817	2	0.00	49,909	99.99
820	1	0.00	49,910	99.99
821	2	0.00	49,912	100.00
822	1	0.00	49,913	100.00
831	1	0.00	49,914	100.00

Table E.17. Scale Score Distribution—Science Grade 11

SS	Freq.	%	Cum. Freq.	Cum. %
650	1,193	3.50	1,193	3.50
651	22	0.06	1,215	3.56
652	31	0.09	1,246	3.65
653	23	0.07	1,269	3.72
654	16	0.05	1,285	3.77
655	14	0.04	1,299	3.81
656	19	0.06	1,318	3.86
657	21	0.06	1,339	3.93
658	19	0.06	1,358	3.98
659	27	0.08	1,385	4.06
660	9	0.03	1,394	4.09
661	22	0.06	1,416	4.15
662	18	0.05	1,434	4.20
663	20	0.06	1,454	4.26
664	23	0.07	1,477	4.33
665	17	0.05	1,494	4.38
666	14	0.04	1,508	4.42
667	18	0.05	1,526	4.47
668	24	0.07	1,550	4.54
669	26	0.08	1,576	4.62
670	35	0.10	1,611	4.72
671	34	0.10	1,645	4.82
672	34	0.10	1,679	4.92
673	19	0.06	1,698	4.98
674	34	0.10	1,732	5.08
675	28	0.08	1,760	5.16
676	41	0.12	1,801	5.28
677	49	0.14	1,850	5.42
678	49	0.14	1,899	5.57
679	50	0.15	1,949	5.71
680	48	0.14	1,997	5.85
681	60	0.18	2,057	6.03
682	66	0.19	2,123	6.22
683	76	0.22	2,199	6.45
684	82	0.24	2,281	6.69
685	80	0.23	2,361	6.92
686	81	0.24	2,442	7.16
687	107	0.31	2,549	7.47
688	82	0.24	2,631	7.71
689	93	0.27	2,724	7.99
690	132	0.39	2,856	8.37
691	124	0.36	2,980	8.74
692	139	0.41	3,119	9.14
693	135	0.40	3,254	9.54
694	133	0.39	3,387	9.93
695	130	0.38	3,517	10.31
696	163	0.48	3,680	10.79
697	168	0.49	3,848	11.28
698	189	0.55	4,037	11.84
699	190	0.56	4,227	12.39
700	193	0.57	4,420	12.96
701	224	0.66	4,644	13.62
702	205	0.60	4,849	14.22
703	243	0.71	5,092	14.93
704	251	0.74	5,343	15.66
705	231	0.68	5,574	16.34
706	303	0.89	5,877	17.23
707	302	0.89	6,179	18.12
708	276	0.81	6,455	18.93
709	340	1.00	6,795	19.92
710	317	0.93	7,112	20.85
711	333	0.98	7,445	21.83
712	385	1.13	7,830	22.96
713	333	0.98	8,163	23.93
714	361	1.06	8,524	24.99
715	418	1.23	8,942	26.22
716	373	1.09	9,315	27.31
717	445	1.30	9,760	28.61
718	416	1.22	10,176	29.83
719	390	1.14	10,566	30.98
720	447	1.31	11,013	32.29
721	452	1.33	11,465	33.61
722	453	1.33	11,918	34.94
723	503	1.47	12,421	36.42
724	467	1.37	12,888	37.79
725	521	1.53	13,409	39.31
726	489	1.43	13,898	40.75
727	511	1.50	14,409	42.25
728	537	1.57	14,946	43.82
729	541	1.59	15,487	45.41
730	527	1.55	16,014	46.95
731	553	1.62	16,567	48.57
732	535	1.57	17,102	50.14
733	555	1.63	17,657	51.77
734	576	1.69	18,233	53.46
735	528	1.55	18,761	55.00
736	547	1.60	19,308	56.61
737	572	1.68	19,880	58.29
738	484	1.42	20,364	59.70
739	550	1.61	20,914	61.32
740	537	1.57	21,451	62.89
741	492	1.44	21,943	64.33
742	487	1.43	22,430	65.76
743	478	1.40	22,908	67.16
744	483	1.42	23,391	68.58
745	453	1.33	23,844	69.91
746	456	1.34	24,300	71.24
747	466	1.37	24,766	72.61
748	442	1.30	25,208	73.91
749	441	1.29	25,649	75.20

SS	Freq.	%	Cum. Freq.	Cum. %
750	444	1.30	26,093	76.50
751	431	1.26	26,524	77.76
752	397	1.16	26,921	78.93
753	380	1.11	27,301	80.04
754	388	1.14	27,689	81.18
755	375	1.10	28,064	82.28
756	396	1.16	28,460	83.44
757	334	0.98	28,794	84.42
758	291	0.85	29,085	85.27
759	314	0.92	29,399	86.19
760	286	0.84	29,685	87.03
761	297	0.87	29,982	87.90
762	266	0.78	30,248	88.68
763	250	0.73	30,498	89.42
764	263	0.77	30,761	90.19
765	260	0.76	31,021	90.95
766	226	0.66	31,247	91.61
767	225	0.66	31,472	92.27
768	201	0.59	31,673	92.86
769	197	0.58	31,870	93.44
770	182	0.53	32,052	93.97
771	174	0.51	32,226	94.48
772	171	0.50	32,397	94.98
773	162	0.47	32,559	95.46
774	138	0.40	32,697	95.86
775	124	0.36	32,821	96.23
776	121	0.35	32,942	96.58
777	122	0.36	33,064	96.94
778	114	0.33	33,178	97.27
779	102	0.30	33,280	97.57
780	77	0.23	33,357	97.80
781	78	0.23	33,435	98.03
782	72	0.21	33,507	98.24
783	69	0.20	33,576	98.44
784	62	0.18	33,638	98.62
785	43	0.13	33,681	98.75
786	51	0.15	33,732	98.90
787	52	0.15	33,784	99.05
788	41	0.12	33,825	99.17
789	33	0.10	33,858	99.27
790	39	0.11	33,897	99.38
791	29	0.09	33,926	99.47
792	26	0.08	33,952	99.54
793	21	0.06	33,973	99.60
794	13	0.04	33,986	99.64
795	17	0.05	34,003	99.69
796	14	0.04	34,017	99.73
797	16	0.05	34,033	99.78
798	6	0.02	34,039	99.80
799	7	0.02	34,046	99.82
800	10	0.03	34,056	99.85
801	5	0.01	34,061	99.86
802	7	0.02	34,068	99.88
803	10	0.03	34,078	99.91
804	2	0.01	34,080	99.92
805	4	0.01	34,084	99.93
806	3	0.01	34,087	99.94
807	4	0.01	34,091	99.95
808	7	0.02	34,098	99.97
809	2	0.01	34,100	99.98
812	2	0.01	34,102	99.98
813	1	0.00	34,103	99.99
816	1	0.00	34,104	99.99
817	3	0.01	34,107	100.00
820	1	0.00	34,108	100.00

Appendix F: Scale Score Distribution Graphs

Figure F.1. Scale Score Distribution—Mathematics Grade 3

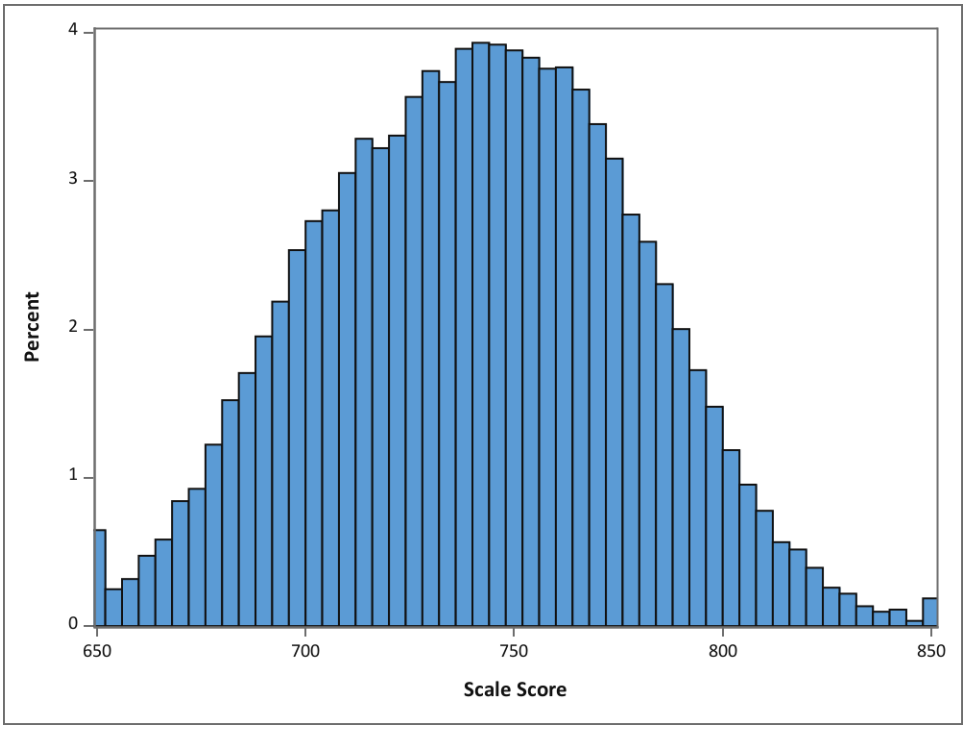


Figure F.2. Scale Score Distribution—Mathematics Grade 4

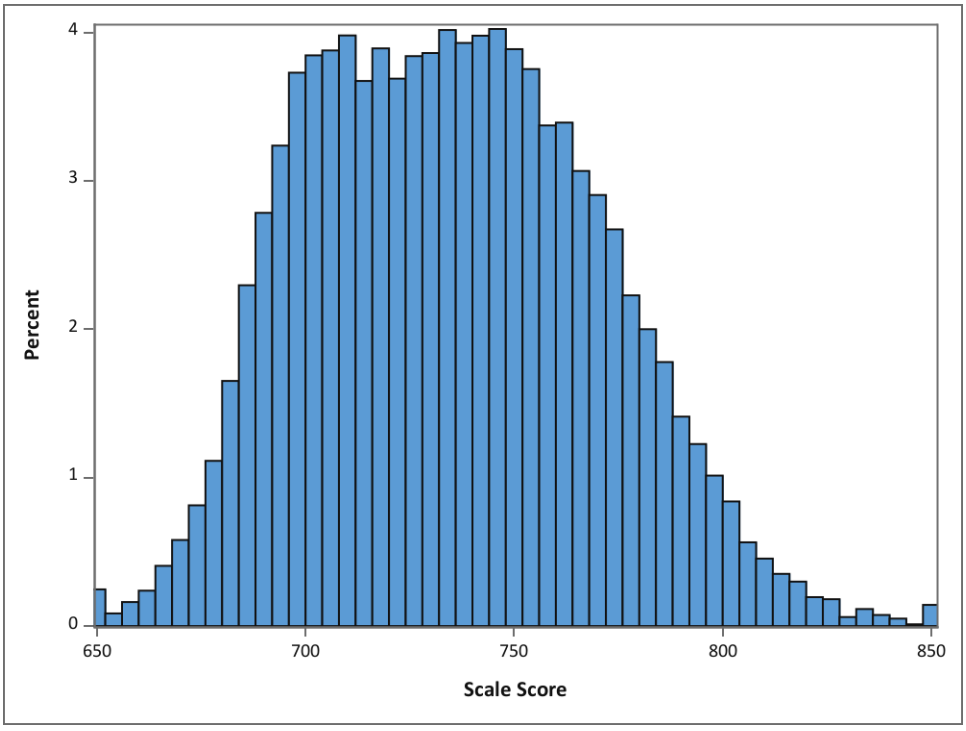


Figure F.3. Scale Score Distribution—Mathematics Grade 5

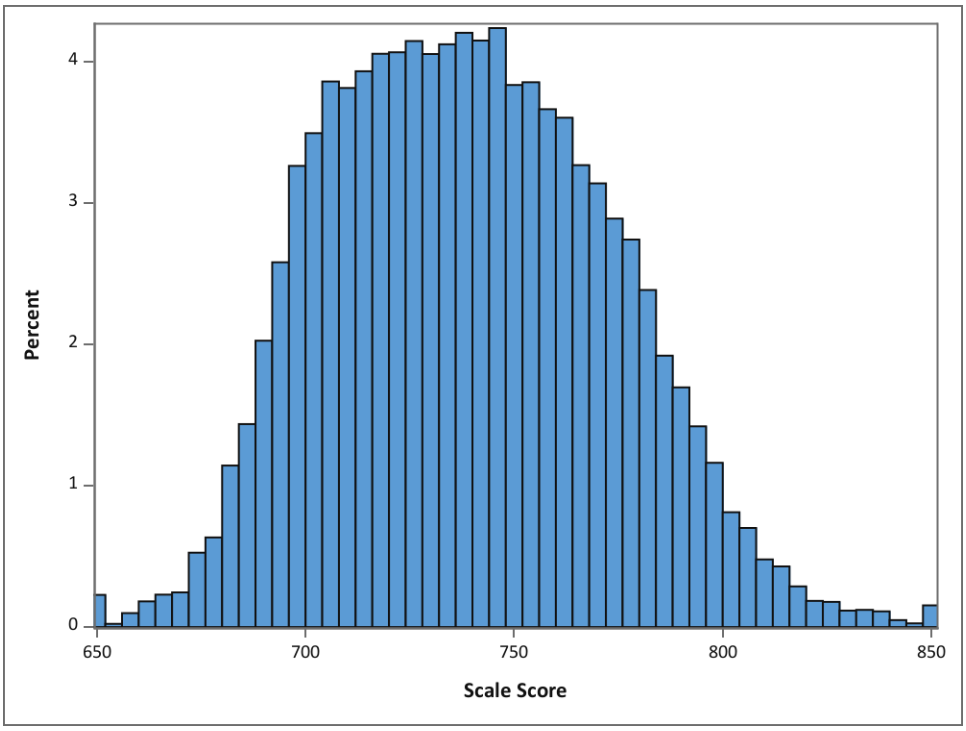


Figure F.4. Scale Score Distribution—Mathematics Grade 6

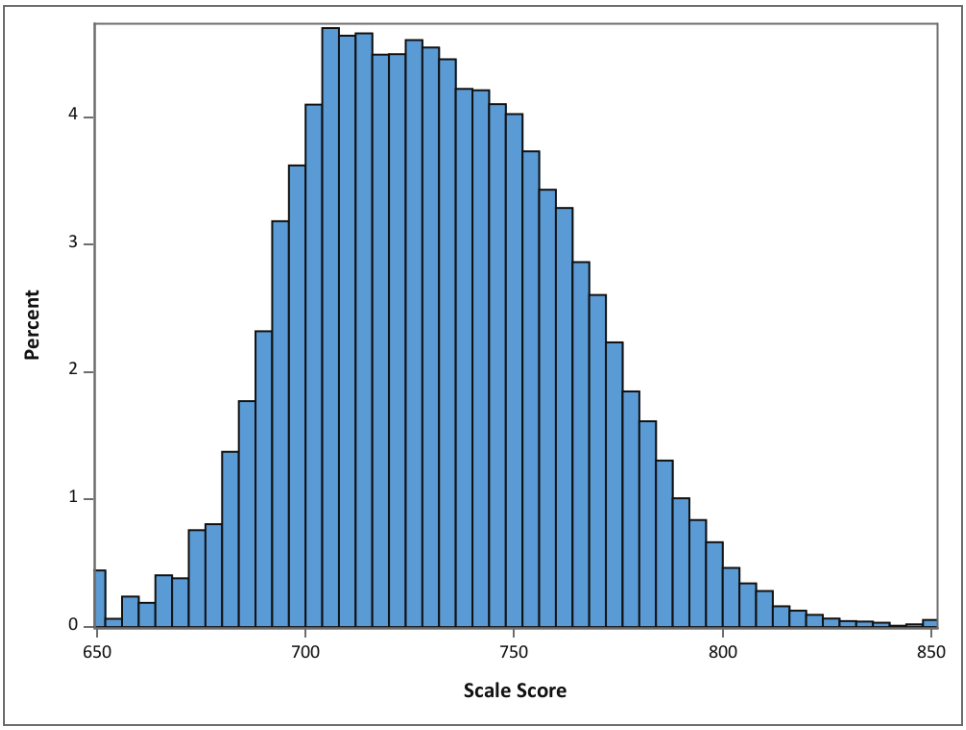


Figure F.5. Scale Score Distribution—Mathematics Grade 7

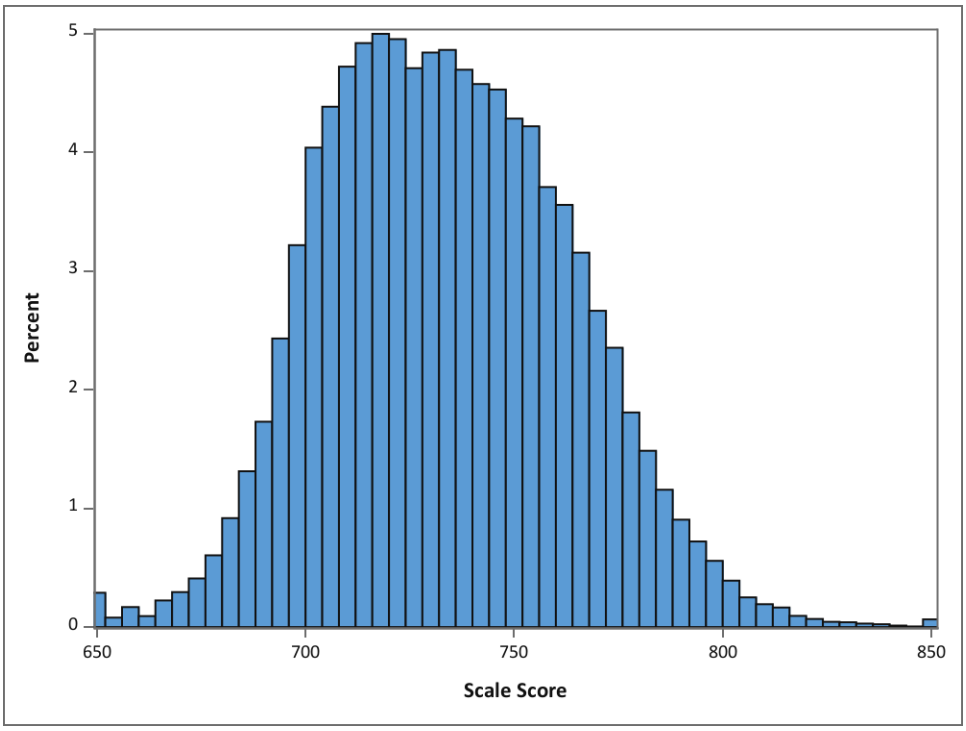


Figure F.6. Scale Score Distribution—Mathematics Grade 8

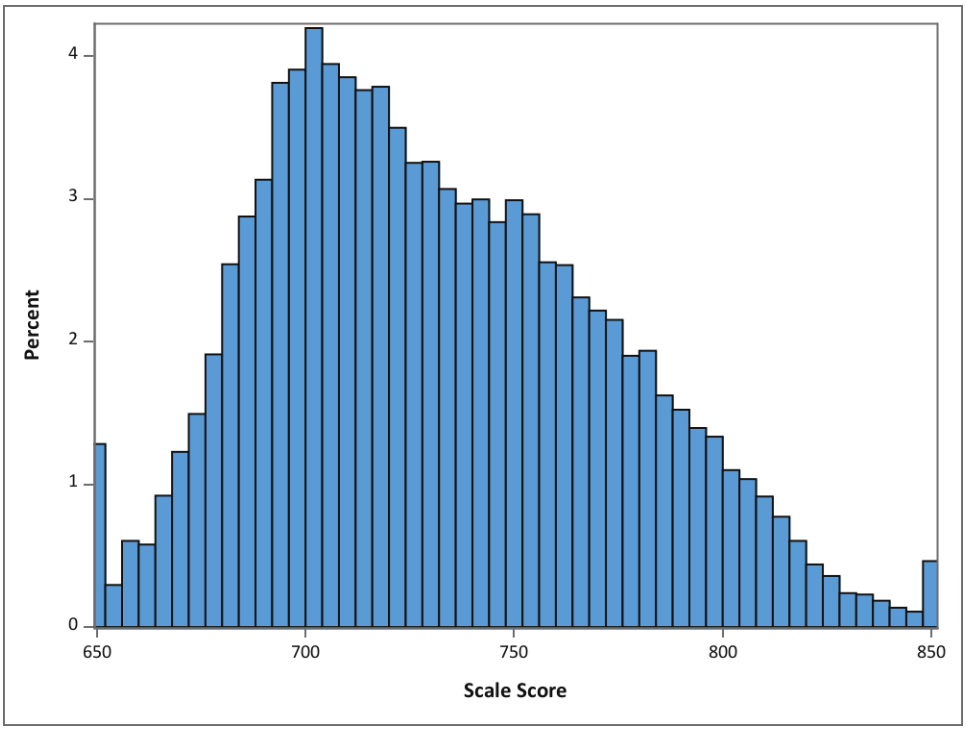


Figure F.7. Scale Score Distribution—ELA Grade 3

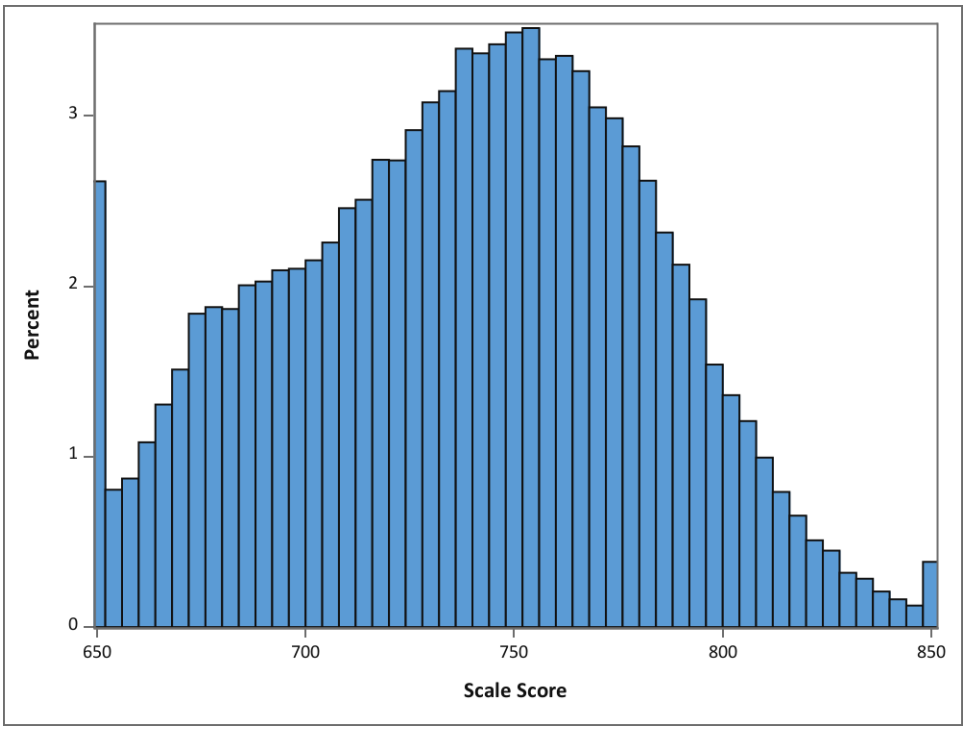


Figure F.8. Scale Score Distribution—ELA Grade 4

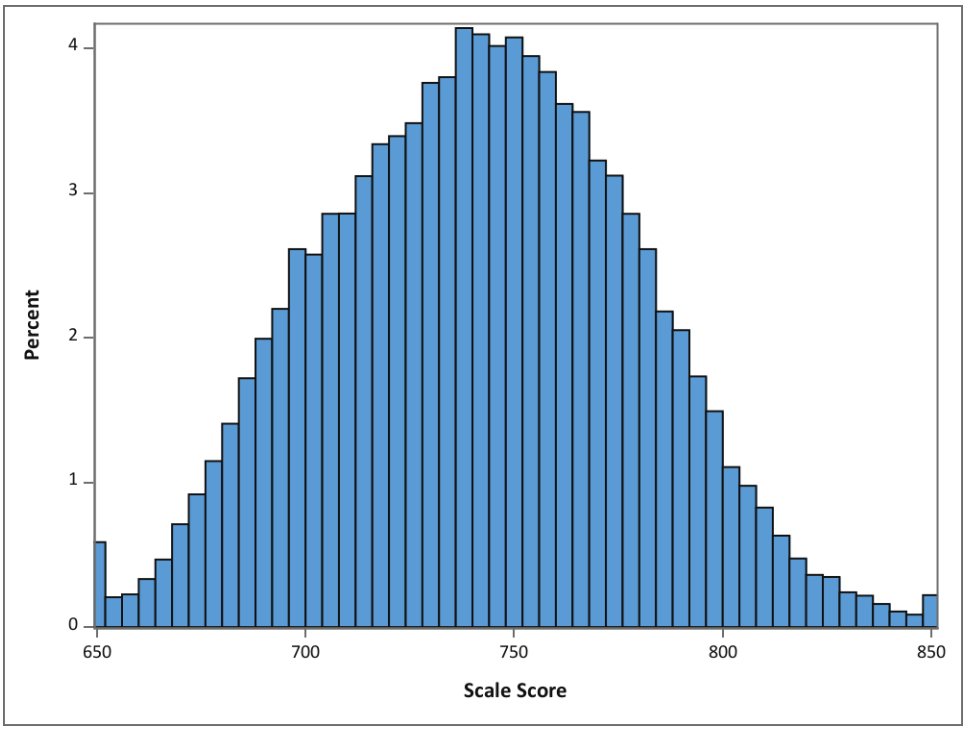


Figure F.9. Scale Score Distribution—ELA Grade 5

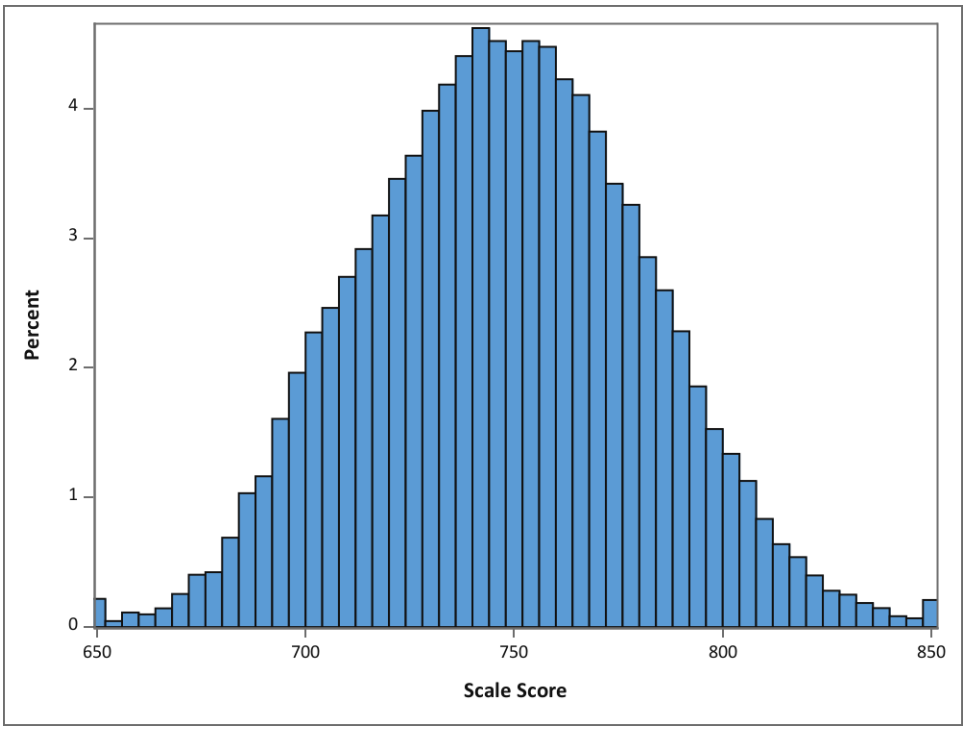


Figure F.10. Scale Score Distribution—ELA Grade 6

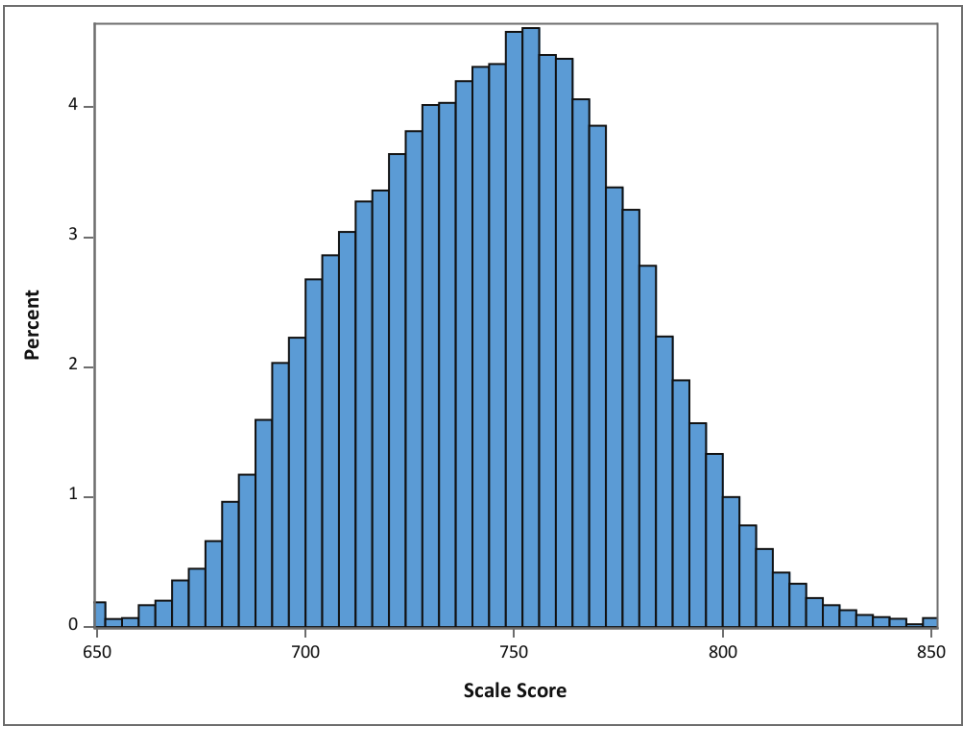


Figure F.11. Scale Score Distribution—ELA Grade 7

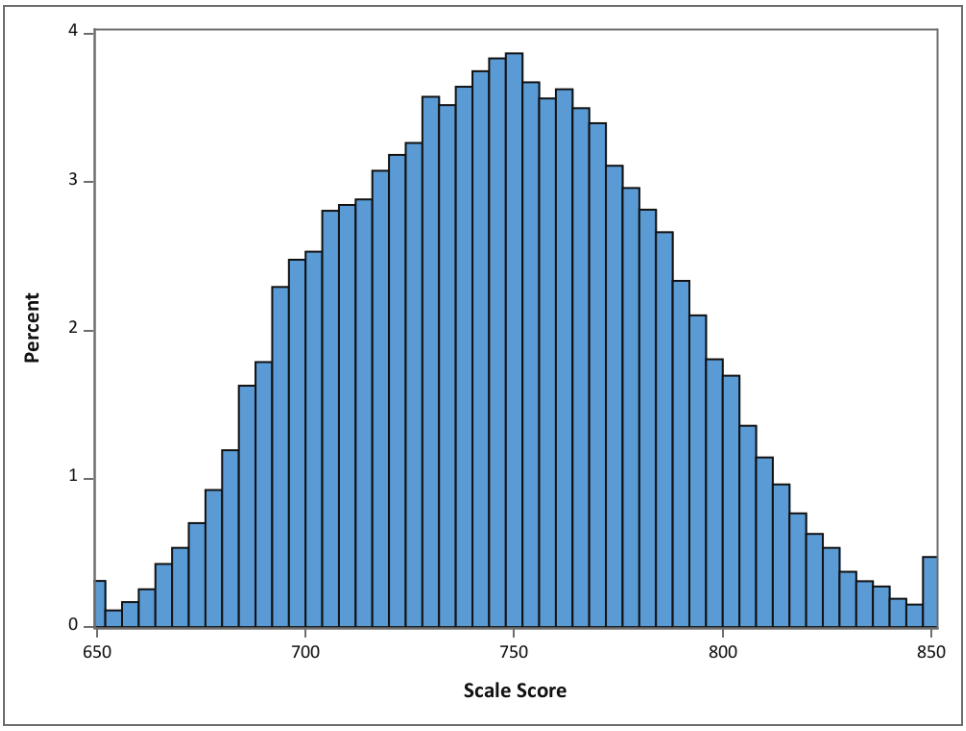


Figure F.12. Scale Score Distribution—ELA Grade 8

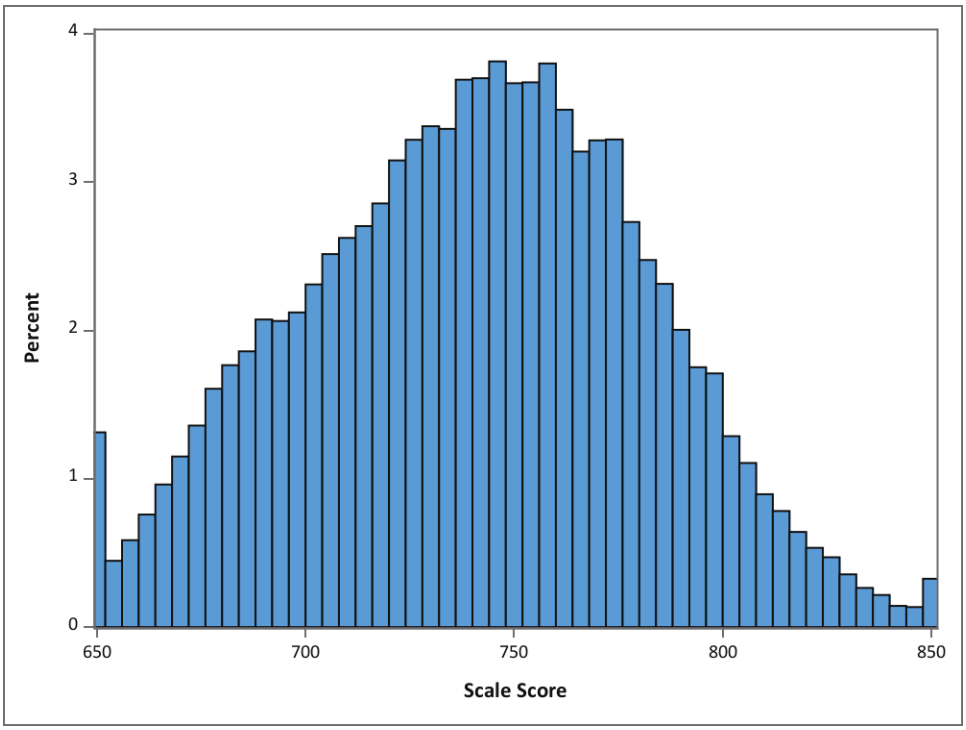


Figure F.13. Scale Score Distribution—CSLA Grade 3

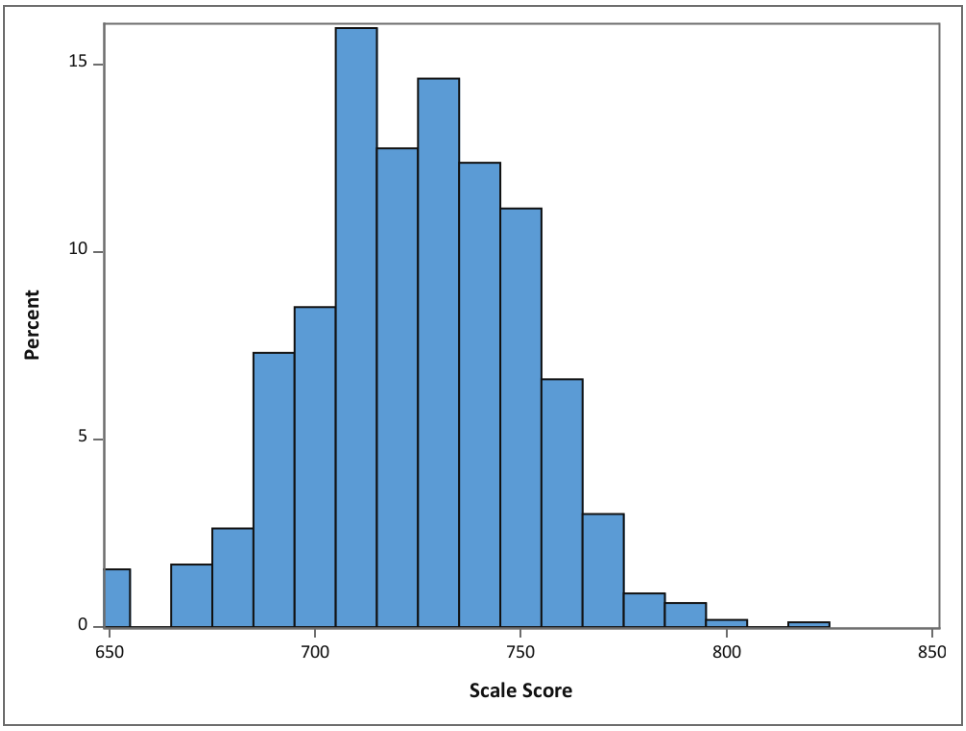


Figure F.14. Scale Score Distribution—CSLA Grade 4

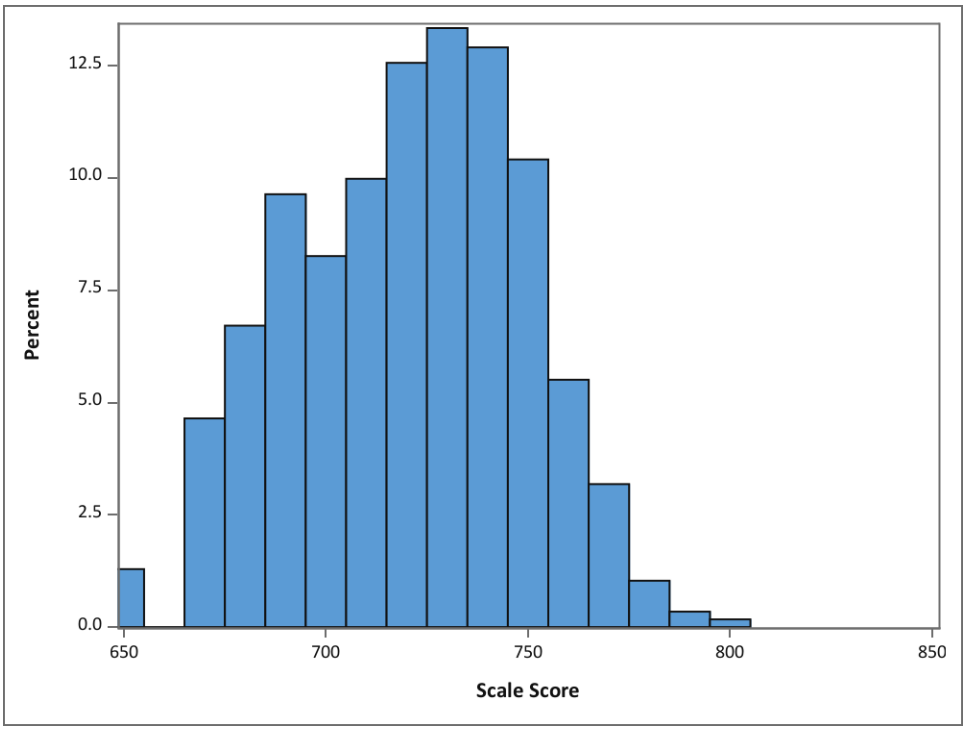


Figure F.15. Scale Score Distribution—Science Grade 5

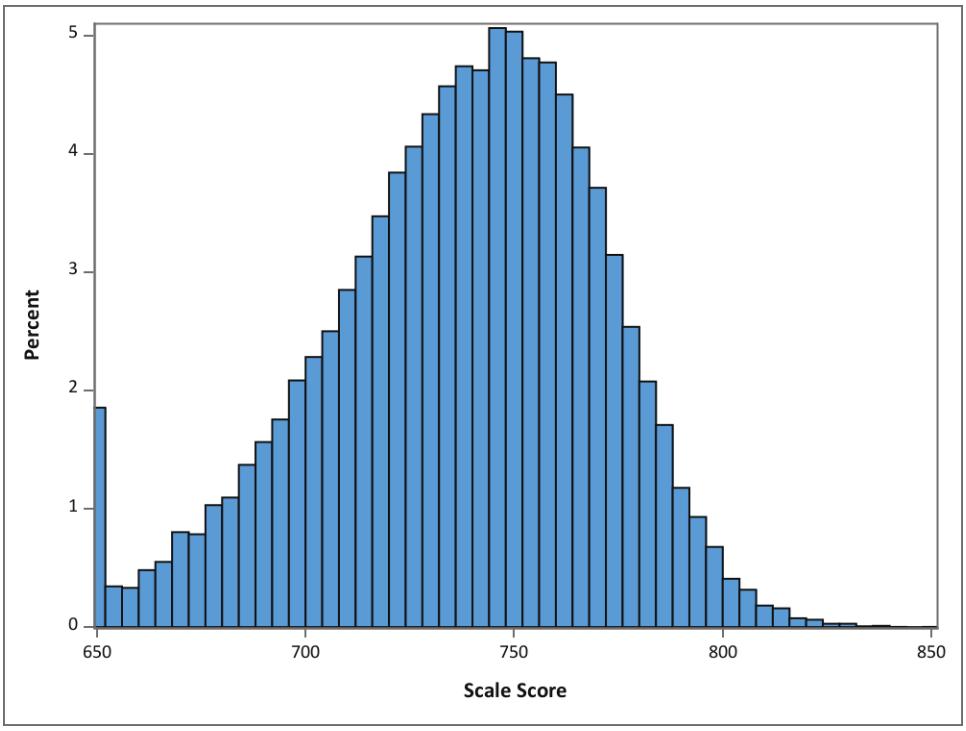


Figure F.16. Scale Score Distribution—Science Grade 8

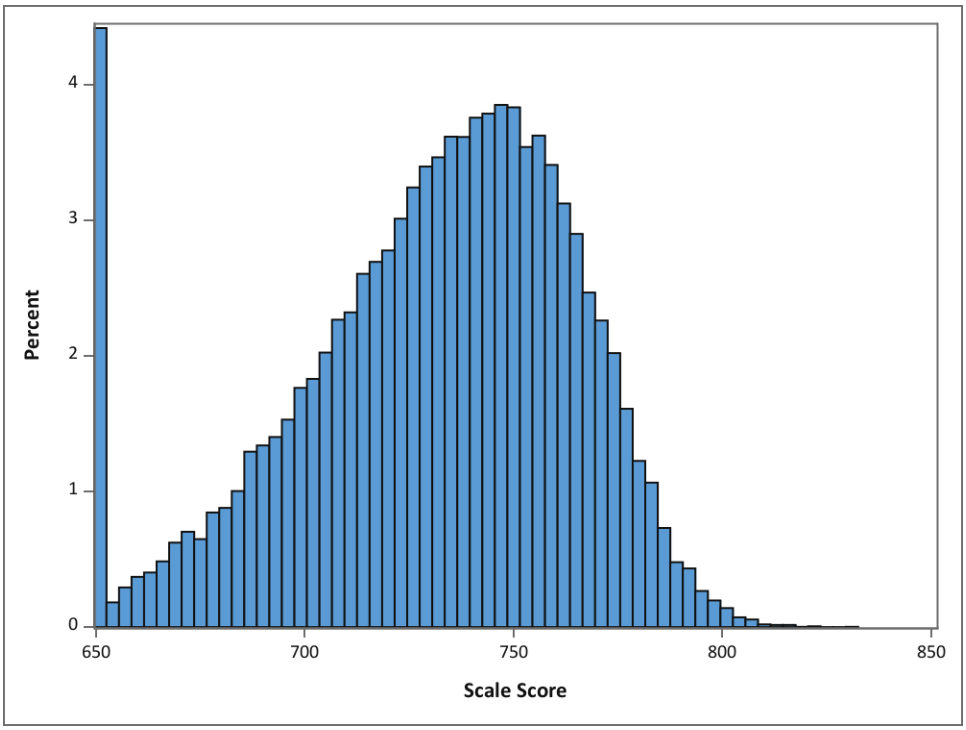
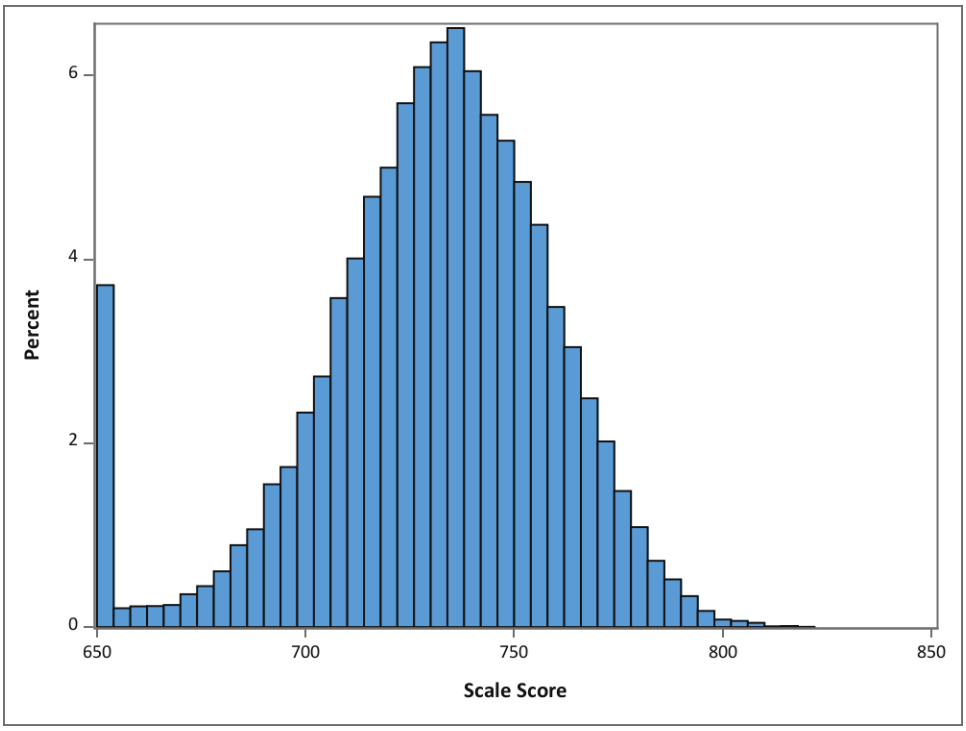


Figure F.17. Scale Score Distribution—Science Grade 11



Appendix G: Scale Score Summary Statistics by Demographic Group

Table G.1. Performance by Subgroup—Mathematics Grade 3

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,508	744.53	36.02	650	850	0.92
IEP	7,221	712.17	35.62	650	850	0.92
No Accommodation	52,386	742.76	36.87	650	850	0.92
Accommodation	4,343	712.06	33.90	650	850	0.91
Am. Indian/Alaska Native	344	724.19	34.88	650	826	0.92
Asian	1,958	757.11	37.86	650	850	0.92
Black	2,605	723.55	35.05	650	850	0.91
Hispanic	19,829	723.53	34.40	650	850	0.91
White	28,527	752.13	34.58	650	850	0.91
Hawaiian/Pacific Islander	187	721.07	34.93	650	822	0.91
Two or More Races	3,104	745.79	37.75	650	850	0.92
Missing	175	764.00	27.00	684	838	0.88
No Economic Disadvantage	31,978	752.52	35.63	650	850	0.91
Economic Disadvantage	24,751	724.77	34.02	650	850	0.91
Female	28,056	738.27	36.66	650	850	0.92
Male	28,663	742.50	38.28	650	850	0.92
Nonbinary	*	*	*	*	*	*
Language Proficiency NA	46,010	745.60	36.43	650	850	0.92
Language Proficiency NEP	3,370	698.36	28.02	650	850	0.87
Language Proficiency LEP	6,328	722.04	29.91	650	850	0.89
Language Proficiency FEP	1,021	759.25	30.26	654	850	0.89
Not Migrant	56,518	740.50	37.54	650	850	0.92
Migrant	211	715.03	29.49	650	788	0.89

*n-count less than 16

Table G.2. Performance by Subgroup—Mathematics Grade 4

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,876	738.61	33.84	650	850	0.91
IEP	7,519	708.95	29.00	650	842	0.90
No Accommodation	52,526	736.89	34.35	650	850	0.92
Accommodation	4,869	711.36	29.70	650	850	0.90
Am. Indian/Alaska Native	350	714.48	30.64	650	809	0.90
Asian	2,025	751.01	37.77	650	850	0.92
Black	2,636	718.41	29.87	650	837	0.90
Hispanic	19,691	718.97	30.09	650	850	0.90
White	29,061	745.21	32.99	650	850	0.91
Hawaiian/Pacific Islander	210	718.35	32.27	650	817	0.91
Two or More Races	3,239	741.31	35.16	650	850	0.92
Missing	183	761.00	28.00	687	850	0.88
No Economic Disadvantage	32,536	745.90	33.77	650	850	0.91
Economic Disadvantage	24,859	720.10	30.21	650	850	0.90

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Female	28,216	732.04	33.30	650	850	0.91
Male	29,164	737.32	35.84	650	850	0.92
Nonbinary	*	*	*	*	*	*
Language Proficiency NA	46,679	739.03	34.20	650	850	0.91
Language Proficiency NEP	2,767	697.85	21.66	650	850	0.80
Language Proficiency LEP	5,679	712.51	24.66	650	824	0.85
Language Proficiency FEP	2,270	746.69	30.27	650	850	0.90
Not Migrant	57,196	734.83	34.70	650	850	0.92
Migrant	199	705.79	25.66	650	810	0.86

*n-count less than 16

Table G.3. Performance by Subgroup—Mathematics Grade 5

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,240	742.63	32.74	650	850	0.92
IEP	7,361	712.00	27.17	650	839	0.90
No Accommodation	51,567	740.85	33.28	650	850	0.92
Accommodation	5,034	716.04	29.18	650	850	0.91
Am. Indian/Alaska Native	378	721.74	29.14	650	840	0.91
Asian	2,108	756.68	35.45	650	850	0.92
Black	2,559	723.99	29.53	650	836	0.91
Hispanic	19,900	723.87	29.29	650	850	0.91
White	28,214	748.56	32.28	650	850	0.91
Hawaiian/Pacific Islander	214	722.74	28.85	650	791	0.90
Two or More Races	3,050	745.20	34.04	650	850	0.92
Missing	178	759.00	25.00	700	833	0.88
No Economic Disadvantage	32,240	749.30	32.99	650	850	0.92
Economic Disadvantage	24,361	724.55	29.09	650	850	0.91
Female	27,712	736.61	33.01	650	850	0.92
Male	28,878	740.59	34.19	650	850	0.93
Nonbinary	*	*	*	*	*	*
Language Proficiency NA	45,854	742.79	33.41	650	850	0.92
Language Proficiency NEP	1,916	703.82	21.32	650	792	0.83
Language Proficiency LEP	5,204	713.31	22.82	650	825	0.86
Language Proficiency FEP	3,627	740.94	28.64	650	850	0.90
Not Migrant	56,391	738.74	33.68	650	850	0.92
Migrant	210	713.63	23.79	650	789	0.87

*n-count less than 16

Table G.4. Performance by Subgroup—Mathematics Grade 6

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	48,555	735.17	30.84	650	850	0.91
IEP	6,569	706.15	24.45	650	850	0.87
No Accommodation	50,434	733.61	31.30	650	850	0.91
Accommodation	4,690	711.35	27.11	650	825	0.89
Am. Indian/Alaska Native	338	716.48	28.25	650	817	0.90
Asian	1,954	749.96	33.93	650	850	0.92
Black	2,400	717.17	28.13	650	823	0.89
Hispanic	19,561	717.47	27.03	650	850	0.88
White	27,683	741.41	30.20	650	850	0.90
Hawaiian/Pacific Islander	196	717.13	25.17	650	778	0.86
Two or More Races	2,856	736.54	31.38	650	850	0.91
Missing	136	758.00	25.00	685	838	0.87
No Economic Disadvantage	31,658	741.69	30.94	650	850	0.91
Economic Disadvantage	23,466	718.25	27.13	650	850	0.88
Female	26,651	729.83	30.78	650	850	0.90
Male	28,456	733.47	32.22	650	850	0.91
Nonbinary	17	747.76	29.46	694	796	0.90
Language Proficiency NA	44,616	735.97	31.17	650	850	0.91
Language Proficiency NEP	1,679	697.07	20.17	650	797	0.76
Language Proficiency LEP	4,124	705.08	21.02	650	810	0.79
Language Proficiency FEP	4,705	727.01	26.40	650	850	0.88
Not Migrant	54,932	731.80	31.58	650	850	0.91
Migrant	192	707.79	23.53	650	782	0.85

Table G.5. Performance by Subgroup—Mathematics Grade 7

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	47,852	736.47	28.85	650	850	0.90
IEP	6,050	709.06	22.23	650	848	0.84
No Accommodation	49,322	735.19	29.21	650	850	0.90
Accommodation	4,580	714.09	25.24	650	841	0.88
Am. Indian/Alaska Native	333	718.63	25.76	650	824	0.87
Asian	1,903	752.56	32.66	650	850	0.92
Black	2,340	720.29	27.43	650	822	0.89
Hispanic	19,498	720.56	25.57	650	850	0.87
White	26,699	742.16	27.87	650	850	0.90
Hawaiian/Pacific Islander	204	720.19	27.03	653	791	0.89
Two or More Races	2,780	738.85	30.07	650	850	0.91
Missing	145	753.00	24.00	681	844	0.89
No Economic Disadvantage	31,222	742.50	28.80	650	850	0.90
Economic Disadvantage	22,680	720.87	25.56	650	850	0.87
Female	25,993	732.58	29.56	650	850	0.90
Male	27,892	734.14	29.40	650	850	0.91
Nonbinary	17	754.24	25.33	706	809	0.89

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Language Proficiency NA	43,776	737.47	29.06	650	850	0.90
Language Proficiency NEP	1,823	701.36	17.85	650	794	0.60
Language Proficiency LEP	3,969	709.75	20.09	650	813	0.77
Language Proficiency FEP	4,334	727.33	25.14	650	848	0.87
Not Migrant	53,736	733.46	29.49	650	850	0.91
Migrant	166	713.06	19.53	668	770	0.79

Table G.6. Performance by Subgroup—Mathematics Grade 8

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	45,360	735.36	40.98	650	850	0.91
IEP	5,443	699.35	27.87	650	850	0.82
No Accommodation	46,544	733.75	41.19	650	850	0.91
Accommodation	4,259	706.86	34.05	650	850	0.88
Am. Indian/Alaska Native	331	712.17	33.93	650	850	0.87
Asian	1,665	758.53	45.91	650	850	0.92
Black	2,221	714.47	35.43	650	850	0.88
Hispanic	19,153	712.93	33.99	650	850	0.87
White	24,672	745.00	40.03	650	850	0.90
Hawaiian/Pacific Islander	171	710.34	32.46	650	802	0.86
Two or More Races	2,469	740.52	43.18	650	850	0.91
Missing	121	757.00	36.00	650	845	0.88
No Economic Disadvantage	29,028	745.04	41.21	650	850	0.91
Economic Disadvantage	21,775	713.45	33.89	650	850	0.87
Female	24,277	730.35	40.44	650	850	0.90
Male	26,512	732.54	42.08	650	850	0.91
Nonbinary	*	*	*	*	*	*
Language Proficiency NA	41,801	737.51	41.03	650	850	0.91
Language Proficiency NEP	1,851	690.02	21.68	650	806	0.60
Language Proficiency LEP	3,784	697.94	24.15	650	826	0.71
Language Proficiency FEP	3,367	717.42	32.90	650	850	0.87
Not Migrant	50,625	731.61	41.31	650	850	0.91
Migrant	178	701.16	30.40	650	808	0.84

*n-count less than 16

Table G.7. Performance by Subgroup—ELA Grade 3

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	47,599	743.24	41.83	650	850	0.90
IEP	6,978	699.32	39.67	650	850	0.92
No Accommodation	50,853	740.22	43.27	650	850	0.91
Accommodation	3,724	702.15	39.40	650	850	0.91
Am. Indian/Alaska Native	344	718.46	42.45	650	850	0.91
Asian	1,931	748.85	44.17	650	850	0.91
Black	2,575	720.40	41.68	650	850	0.91
Hispanic	17,888	718.95	41.43	650	850	0.91
White	28,376	749.70	41.21	650	850	0.90
Hawaiian/Pacific Islander	189	713.42	41.38	650	808	0.91
Two or More Races	3,101	744.16	43.42	650	850	0.91
Missing	173	766.56	34.41	664	836	0.86
No Economic Disadvantage	31,345	751.29	41.19	650	850	0.89
Economic Disadvantage	23,232	719.18	41.00	650	850	0.91
Female	27,032	740.90	44.25	650	850	0.91
Male	27,535	734.40	43.65	650	850	0.91
Nonbinary	*	*	*	*	*	*
Language Proficiency NA	45,930	742.81	42.84	650	850	0.91
Language Proficiency NEP	2,046	678.47	25.48	650	793	0.84
Language Proficiency LEP	5,580	712.55	35.24	650	841	0.89
Language Proficiency FEP	1,021	759.93	31.67	653	850	0.83
Not Migrant	54,395	737.76	44.03	650	850	0.91
Migrant	182	697.48	35.77	650	817	0.90

*n-count less than 16

Table G.8. Performance by Subgroup—ELA Grade 4

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	48,361	746.49	35.24	650	850	0.87
IEP	7,349	707.13	32.22	650	846	0.88
No Accommodation	51,350	743.73	36.60	650	850	0.88
Accommodation	4,360	712.64	33.51	650	850	0.88
Am. Indian/Alaska Native	348	722.35	33.65	650	807	0.87
Asian	1,989	749.68	38.38	650	850	0.88
Black	2,619	723.97	33.32	650	850	0.87
Hispanic	18,144	724.32	33.36	650	850	0.87
White	28,989	752.25	35.36	650	850	0.87
Hawaiian/Pacific Islander	208	722.29	32.79	650	850	0.87
Two or More Races	3,233	748.96	36.85	650	850	0.88
Missing	180	769.00	26.00	704	835	0.77
No Economic Disadvantage	32,059	753.28	35.55	650	850	0.87
Economic Disadvantage	23,651	725.06	33.28	650	850	0.87
Female	27,410	743.96	37.62	650	850	0.88
Male	28,285	738.70	36.82	650	850	0.89
Nonbinary	*	*	*	*	*	*

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Language Proficiency NA	46,664	745.80	36.52	650	850	0.88
Language Proficiency NEP	1,645	689.41	20.96	650	815	0.77
Language Proficiency LEP	5,130	712.78	24.85	650	827	0.81
Language Proficiency FEP	2,271	750.86	27.16	650	850	0.81
Not Migrant	55,537	741.41	37.28	650	850	0.88
Migrant	173	705.83	31.35	650	796	0.85

*n-count less than 16

Table G.9. Performance by Subgroup—ELA Grade 5

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	48,625	752.00	31.80	650	850	0.88
IEP	7,328	714.69	28.13	650	850	0.88
No Accommodation	51,186	749.59	33.03	650	850	0.89
Accommodation	4,767	720.56	30.12	650	850	0.89
Am. Indian/Alaska Native	380	731.90	28.81	650	807	0.86
Asian	2,070	758.71	33.66	657	850	0.89
Black	2,536	733.38	30.90	650	850	0.88
Hispanic	19,344	732.53	30.95	650	850	0.88
White	28,169	756.98	31.79	650	850	0.88
Hawaiian/Pacific Islander	215	731.35	32.55	667	839	0.88
Two or More Races	3,059	753.78	33.44	650	850	0.89
Missing	180	769.00	30.00	680	850	0.82
No Economic Disadvantage	31,950	757.80	32.26	650	850	0.88
Economic Disadvantage	24,003	732.89	30.33	650	850	0.88
Female	27,411	751.48	34.03	650	850	0.89
Male	28,530	742.91	33.00	650	850	0.89
Nonbinary	*	*	*	*	*	*
Language Proficiency NA	45,890	751.71	32.91	650	850	0.89
Language Proficiency NEP	1,237	698.75	20.82	650	803	0.81
Language Proficiency LEP	5,192	716.77	22.79	650	814	0.82
Language Proficiency FEP	3,634	748.98	24.95	650	850	0.84
Not Migrant	55,763	747.20	33.76	650	850	0.89
Migrant	190	721.43	28.64	650	805	0.87

*n-count less than 16

Table G.10. Performance by Subgroup—ELA Grade 6

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	48,007	747.50	31.44	650	850	0.89
IEP	6,555	711.42	27.00	650	825	0.89
No Accommodation	50,130	745.42	32.39	650	850	0.90
Accommodation	4,432	717.59	29.92	650	829	0.90
Am. Indian/Alaska Native	334	727.68	30.90	657	822	0.90
Asian	1,934	757.00	33.72	650	850	0.90
Black	2,380	730.83	31.77	650	835	0.90
Hispanic	19,077	729.11	30.39	650	850	0.89
White	27,638	752.57	31.07	650	850	0.89
Hawaiian/Pacific Islander	193	728.72	29.32	650	823	0.89
Two or More Races	2,868	748.60	32.62	650	850	0.90
Missing	138	765.00	23.00	700	832	0.83
No Economic Disadvantage	31,408	753.32	31.42	650	850	0.89
Economic Disadvantage	23,154	729.39	30.18	650	850	0.89
Female	26,393	746.86	33.19	650	850	0.90
Male	28,151	739.68	32.60	650	850	0.91
Nonbinary	18	761.72	27.48	720	807	0.87
Language Proficiency NA	44,640	747.56	32.34	650	850	0.90
Language Proficiency NEP	1,104	696.61	20.12	650	807	0.83
Language Proficiency LEP	4,106	711.22	21.59	650	806	0.83
Language Proficiency FEP	4,712	740.22	26.09	650	850	0.87
Not Migrant	54,382	743.25	33.06	650	850	0.90
Migrant	180	715.32	27.31	662	801	0.89

Table G.11. Performance by Subgroup—ELA Grade 7

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	47,266	750.60	37.49	650	850	0.91
IEP	6,068	709.82	28.93	650	850	0.88
No Accommodation	48,985	748.53	38.22	650	850	0.91
Accommodation	4,349	716.95	33.63	650	850	0.90
Am. Indian/Alaska Native	332	728.31	35.82	650	828	0.91
Asian	1,880	763.62	39.47	650	850	0.91
Black	2,335	732.96	35.66	650	849	0.89
Hispanic	18,970	729.82	35.27	650	850	0.90
White	26,690	756.78	37.04	650	850	0.90
Hawaiian/Pacific Islander	202	731.01	37.76	650	850	0.90
Two or More Races	2,781	753.26	38.47	650	850	0.91
Missing	144	769.00	28.00	664	824	0.88
No Economic Disadvantage	30,973	757.33	37.60	650	850	0.90
Economic Disadvantage	22,361	730.20	34.83	650	850	0.90
Female	25,739	752.04	38.76	650	850	0.91
Male	27,578	740.27	38.05	650	850	0.91
Nonbinary	17	777.12	28.01	716	829	0.91

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Language Proficiency NA	43,829	751.36	38.01	650	850	0.91
Language Proficiency NEP	1,238	690.87	20.31	650	785	0.81
Language Proficiency LEP	3,944	711.17	24.36	650	809	0.85
Language Proficiency FEP	4,323	738.73	31.38	650	850	0.89
Not Migrant	53,185	746.05	38.83	650	850	0.91
Migrant	149	713.52	29.95	650	809	0.88

Table G.12. Performance by Subgroup—ELA Grade 8

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	44,883	745.01	39.64	650	850	0.89
IEP	5,458	700.86	31.50	650	844	0.88
No Accommodation	46,328	742.82	40.46	650	850	0.89
Accommodation	4,013	710.18	37.60	650	850	0.90
Am. Indian/Alaska Native	331	723.15	36.20	650	850	0.89
Asian	1,641	761.44	42.42	650	850	0.89
Black	2,221	725.64	39.46	650	850	0.89
Hispanic	18,741	723.07	38.16	650	850	0.89
White	24,648	752.47	38.26	650	850	0.88
Hawaiian/Pacific Islander	169	722.21	34.07	650	801	0.87
Two or More Races	2,461	749.72	40.55	650	850	0.89
Missing	129	759.00	32.00	658	834	0.80
No Economic Disadvantage	28,808	752.71	39.22	650	850	0.89
Economic Disadvantage	21,533	723.51	37.71	650	850	0.89
Female	24,083	747.03	41.21	650	850	0.89
Male	26,244	733.95	40.17	650	850	0.90
Nonbinary	*	*	*	*	*	*
Language Proficiency NA	41,849	746.48	39.68	650	850	0.89
Language Proficiency NEP	1,340	679.85	21.26	650	771	0.80
Language Proficiency LEP	3,783	701.94	27.24	650	823	0.84
Language Proficiency FEP	3,369	729.44	33.07	650	850	0.87
Not Migrant	50,181	740.33	41.17	650	850	0.90
Migrant	160	704.91	35.01	650	802	0.89

*n-count less than 16

Table G.13. Performance by Subgroup—CSLA Grade 3

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	1,391	726.29	25.60	650	823	0.86
IEP	169	700.15	23.02	650	785	0.83
No Accommodation	1,345	724.53	26.23	650	823	0.86
Accommodation	215	716.74	27.91	650	800	0.87
Am. Indian/Alaska Native	*	*	*	*	*	*
Asian	*	*	*	*	*	*
Black	*	*	*	*	*	*
Hispanic	1,550	723.43	26.57	650	823	0.86
White	10	727.20	32.17	686	793	0.90
Hawaiian/Pacific Islander	*	*	*	*	*	*
Two or More Races	*	*	*	*	*	*
Missing	*	*	*	*	*	*
No Economic Disadvantage	336	719.05	28.44	650	823	0.87
Economic Disadvantage	1,224	724.67	25.95	650	800	0.86
Female	762	726.34	26.35	650	823	0.86
Male	797	720.68	26.56	650	800	0.86
Nonbinary	*	*	*	*	*	*
Language Proficiency NA	*	*	*	*	*	*
Language Proficiency NEP	817	716.13	26.26	650	804	0.86
Language Proficiency LEP	743	731.52	24.57	650	823	0.85
Language Proficiency FEP	*	*	*	*	*	*
Not Migrant	1,543	723.42	26.64	650	823	0.86
Migrant	17	727.41	21.88	686	764	0.88

*n-count less than 16

Table G.14. Performance by Subgroup—CSLA Grade 4

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	1,029	723.25	27.21	650	804	0.87
IEP	133	693.69	21.87	650	760	0.83
No Accommodation	989	721.75	27.82	650	804	0.87
Accommodation	173	709.10	28.41	650	783	0.89
Am. Indian/Alaska Native	*	*	*	*	*	*
Asian	*	*	*	*	*	*
Black	*	*	*	*	*	*
Hispanic	1,154	719.89	28.24	650	804	0.88
White	*	*	*	*	*	*
Hawaiian/Pacific Islander	*	*	*	*	*	*
Two or More Races	*	*	*	*	*	*
Missing	*	*	*	*	*	*
No Economic Disadvantage	241	715.45	27.36	650	800	0.88
Economic Disadvantage	921	721.03	28.39	650	804	0.88
Female	582	724.29	28.01	650	800	0.87
Male	580	715.43	27.83	650	804	0.88
Nonbinary	*	*	*	*	*	*

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Language Proficiency NA	*	*	*	*	*	*
Language Proficiency NEP	626	713.23	28.62	650	804	0.88
Language Proficiency LEP	536	727.62	25.78	665	800	0.86
Language Proficiency FEP	*	*	*	*	*	*
Not Migrant	1,152	719.83	28.28	650	804	0.88
Migrant	*	*	*	*	*	*

*n-count less than 16

Table G.15. Performance by Subgroup—Science Grade 5

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	48,828	740.93	30.74	650	850	0.90
IEP	7,257	705.76	32.83	650	825	0.88
No Accommodation	51,523	738.62	32.19	650	850	0.90
Accommodation	4,562	711.11	33.84	650	828	0.89
Am. Indian/Alaska Native	374	720.29	31.52	650	795	0.88
Asian	2,094	747.56	33.60	650	839	0.91
Black	2,529	720.45	32.94	650	814	0.89
Hispanic	19,709	721.26	32.16	650	832	0.88
White	27,973	747.11	29.02	650	841	0.89
Hawaiian/Pacific Islander	210	718.96	32.42	650	783	0.87
Two or More Races	3,017	743.29	31.67	650	850	0.90
Missing	179	756.00	24.00	696	834	0.85
No Economic Disadvantage	31,999	746.90	30.28	650	850	0.90
Economic Disadvantage	24,086	722.41	31.70	650	838	0.88
Female	27,441	736.74	32.17	650	850	0.90
Male	28,635	736.02	34.14	650	841	0.91
Nonbinary	*	*	*	*	*	*
Language Proficiency NA	45,449	741.47	31.23	650	850	0.90
Language Proficiency NEP	1,869	687.37	28.31	650	791	0.79
Language Proficiency LEP	5,158	706.90	26.41	650	813	0.79
Language Proficiency FEP	3,609	739.82	24.44	650	838	0.85
Not Migrant	55,876	736.49	33.14	650	850	0.90
Migrant	209	706.72	32.06	650	787	0.86

*n-count less than 16

Table G.16. Performance by Subgroup—Science Grade 8

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	44,581	734.20	32.18	650	831	0.91
IEP	5,333	700.61	32.10	650	801	0.86
No Accommodation	46,136	732.51	33.00	650	831	0.91
Accommodation	3,778	707.44	34.84	650	811	0.90
Am. Indian/Alaska Native	324	716.42	32.13	650	794	0.89
Asian	1,653	744.69	31.83	650	814	0.92
Black	2,177	716.40	33.11	650	799	0.89
Hispanic	18,818	715.21	33.06	650	806	0.89
White	24,217	742.31	29.16	650	831	0.91
Hawaiian/Pacific Islander	168	711.54	33.66	650	781	0.88
Two or More Races	2,431	738.75	31.09	650	814	0.91
Missing	126	747.00	26.00	650	794	0.89
No Economic Disadvantage	28,543	741.35	30.39	650	831	0.91
Economic Disadvantage	21,371	716.27	32.80	650	806	0.89
Female	23,824	730.95	32.88	650	821	0.91
Male	26,076	730.29	34.63	650	831	0.92
Nonbinary	*	*	*	*	*	*
Language Proficiency NA	41,054	736.37	31.43	650	831	0.91
Language Proficiency NEP	1,803	681.57	27.96	650	765	0.72
Language Proficiency LEP	3,735	699.53	27.38	650	793	0.75
Language Proficiency FEP	3,322	720.94	28.22	650	806	0.86
Not Migrant	49,739	730.72	33.76	650	831	0.92
Migrant	175	700.48	34.14	650	780	0.88

*n-count less than 16

Table G.17. Performance by Subgroup—Science Grade 11

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	31,019	732.21	27.99	650	820	0.88
IEP	3,089	707.31	28.96	650	800	0.83
No Accommodation	31,394	731.22	28.32	650	820	0.88
Accommodation	2,714	715.27	32.21	650	807	0.88
Am. Indian/Alaska Native	235	721.73	28.33	650	782	0.87
Asian	970	737.87	30.24	650	804	0.90
Black	1,618	715.60	29.65	650	795	0.85
Hispanic	14,540	719.81	27.30	650	807	0.84
White	15,087	740.41	26.11	650	820	0.87
Hawaiian/Pacific Islander	115	717.97	30.16	650	778	0.87
Two or More Races	1,459	734.05	28.53	650	817	0.88
Missing	84	762.00	23.00	682	812	0.87
No Economic Disadvantage	19,199	737.13	27.96	650	820	0.88
Economic Disadvantage	14,909	720.70	27.61	650	807	0.85
Female	16,007	729.32	26.99	650	817	0.87
Male	18,064	730.47	30.61	650	820	0.89
Nonbinary	37	750.08	21.09	703	808	0.89

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Language Proficiency NA	29,284	734.17	27.09	650	820	0.87
Language Proficiency NEP	1,023	687.19	26.46	650	751	0.62
Language Proficiency LEP	2,259	702.08	24.22	650	778	0.64
Language Proficiency FEP	1,542	719.08	22.46	650	795	0.76
Not Migrant	33,968	730.05	28.93	650	820	0.88
Migrant	140	706.54	29.07	650	770	0.81

Appendix H: Summary Statistics for Points Earned by Subclaim

Table H.1. Points Earned Summary by Subclaim—Mathematics

Subclaim	Grade	Mean	SD	Min.	Max.	Average % Correct
Subclaim A	3	11.4	5.2	0	22	52.0
	4	11.9	5.8	0	24	49.8
	5	11.7	6.1	0	23	51.0
	6	7.4	4.9	0	20	37.2
	7	7.8	5.2	0	23	33.7
	8	9.7	5.5	0	24	40.5
Subclaim B	3	4.8	2.5	0	9	52.9
	4	2.9	1.9	0	7	41.2
	5	3.5	2.2	0	8	44.5
	6	2.9	2.3	0	10	29.4
	7	3.1	1.9	0	8	39.1
	8	2.5	2.0	0	7	35.8
Subclaim C	3	3.0	2.7	0	11	27.5
	4	3.4	3.0	0	11	30.4
	5	2.3	2.6	0	11	21.2
	6	3.6	2.8	0	11	32.5
	7	3.0	3.1	0	11	27.1
	8	2.2	2.6	0	11	19.8
Subclaim D	3	2.8	2.4	0	9	31.3
	4	2.6	2.4	0	9	28.6
	5	2.4	2.6	0	9	26.2
	6	1.8	2.4	0	9	19.4
	7	2.3	2.4	0	9	25.4
	8	1.3	2.0	0	9	13.9

Table H.2. Points Earned Summary by Subclaim—ELA

Subclaim	Grade	Mean	SD	Min.	Max.	Average % Correct
RL	3	7.8	4.2	0	17	45.8
	4	8.3	4.0	0	20	41.7
	5	8.4	4.3	0	18	46.9
	6	9.2	4.4	0	18	51.1
	7	8.1	4.1	0	18	45.0
	8	8.3	4.2	0	18	46.1
RI	3	7.7	3.8	0	14	55.0
	4	6.5	4.1	0	18	36.0
	5	6.1	3.7	0	18	33.8
	6	9.3	4.5	0	21	42.4
	7	10.3	5.2	0	22	47.0
	8	9.1	4.9	0	22	41.5
RV	3	6.1	2.9	0	10	60.7
	4	4.9	2.2	0	8	62.0
	5	5.3	2.4	0	8	66.0
	6	4.6	2.4	0	8	57.8
	7	5.6	2.9	0	10	55.6
	8	6.1	2.9	0	10	61.0
WE (unweighted)	3	1.1	1.1	0	6	18.5
	4	1.9	1.4	0	7	26.6
	5	2.0	1.4	0	7	29.1
	6	2.2	1.7	0	8	27.5
	7	2.3	1.8	0	8	29.3
	8	2.4	2.0	0	8	30.2
WKL	3	1.2	1.1	0	6	19.6
	4	1.4	1.4	0	6	23.6
	5	1.6	1.4	0	6	26.8
	6	1.8	1.6	0	6	30.1
	7	2.0	1.7	0	6	33.8
	8	2.3	1.9	0	6	38.4

Note. RL = Reading: Literary Text, RI = Reading: Informational Text, RV = Reading: Vocabulary, WE = Writing: Written Expression, WKL = Writing: Knowledge and Use of Language Conventions. Results for WE are unweighted.

Table H.3. Points Earned Summary by Subclaim—CSLA

Subclaim	Grade	Mean	SD	Min.	Max.	Average % Correct
RL	3	4.9	3.7	0	17	28.7
	4	8.4	5.1	0	19	42.0
RI	3	3.6	2.7	0	14	25.7
	4	4.1	3.3	0	16	22.7
RV	3	4.5	2.8	0	10	45.2
	4	3.9	2.3	0	8	49.0
WE (unweighted)	3	1.3	1.5	0	6	21.5
	4	2.0	1.9	0	7	28.2
WKL	3	1.6	1.6	0	6	25.9
	4	1.7	1.6	0	6	27.6

Note. RL = Reading: Literary Text, RI = Reading: Informational Text, RV = Reading: Vocabulary, WE = Writing: Written Expression, WKL = Writing: Knowledge and Use of Language Conventions. Results for WE are unweighted.

Table H.4. Points Earned Summary by Content Standard—Science

Content Standard	Grade	Mean	SD	Min.	Max.	Average % Correct
Physical Science	5	8.0	4.3	0	18	44.3
	8	7.5	3.8	0	21	35.9
	11	6.5	3.4	0	15	43.3
Life Science	5	5.4	2.7	0	12	45.1
	8	8.5	5.2	0	22	38.8
	11	5.1	2.9	0	14	36.1
Earth and Space Science	5	8.9	4.3	0	21	42.2
	8	5.4	3.7	0	18	30.1
	11	3.7	2.6	0	13	28.3

Appendix I: Classical Item-Level Statistics

Table I.1. SR Item Classical Statistics—Mathematics Grade 3

Item	Omit %	*P*-value	Item–Total Correlation
1	3.6	0.48	0.57
2	2.4	0.57	0.60
3	3.9	0.38	0.65
4	0.4	0.25	0.58
5	0.5	0.58	0.54
6	0.4	0.29	0.47
7	0.8	0.24	0.41
8	0.3	0.63	0.49
9	1.0	0.30	0.53
10	0.4	0.46	0.67
11	0.7	0.67	0.59
12	0.2	0.45	0.53
13	0.2	0.78	0.46
14	0.1	0.77	0.50
15	3.5	0.54	0.43
16	0.5	0.63	0.53
17	0.6	0.77	0.50
18	0.3	0.77	0.33
19	1.6	0.52	0.50
20	0.1	0.82	0.46
21	0.2	0.49	0.43
22	0.8	0.62	0.51
23	0.2	0.81	0.40
24	0.1	0.76	0.50
25	0.2	0.59	0.58
26	0.1	0.34	0.58
27	2.4	0.48	0.60
28	2.0	0.28	0.64
29	0.5	0.35	0.73
30	0.5	0.33	0.79

Table I.2. CR Item Classical Statistics—Mathematics Grade 3

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	5%	6%	*P*-value	Item-Total Correlation
1	4	0.5	33.0	26.9	14.1	15.6	10.0	–	–	0.35	0.73
2	6	0.5	15.5	30.1	25.2	7.8	14.2	2.4	4.4	0.33	0.79
3	3	4.9	63.6	8.1	12.8	10.7	–	–	–	0.22	0.70
4	3	2.0	53.2	14.5	15.1	15.2	–	–	–	0.30	0.76
5	4	1.0	31.8	42.2	16.9	6.8	1.3	–	–	0.25	0.64

Table I.3. SR Item Classical Statistics—Mathematics Grade 4

Item	Omit %	*P*-value	Item–Total Correlation
1	0.9	0.23	0.54
2	0.3	0.21	0.57
3	0.7	0.21	0.64
4	0.1	0.66	0.41
5	1.3	0.34	0.62
6	0.2	0.34	0.61
7	2.8	0.42	0.51
8	4.1	0.65	0.29
9	0.5	0.59	0.45
10	0.2	0.43	0.45
11	0.1	0.64	0.52
12	2.2	0.61	0.42
13	0.1	0.74	0.45
14	1.2	0.46	0.51
15	0.1	0.48	0.40
16	2.1	0.46	0.57
17	0.8	0.65	0.56
18	0.1	0.75	0.52
19	0.6	0.27	0.42
20	3.6	0.67	0.55
21	0.3	0.55	0.54
22	0.2	0.55	0.67
23	0.7	0.52	0.64
24	0.5	0.39	0.62
25	1.5	0.24	0.70
26	2.3	0.20	0.61
27	0.3	0.40	0.76
28	3.7	0.30	0.78
29	0.2	0.22	0.79

Table I.4. CR Item Classical Statistics—Mathematics Grade 4

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	5%	6%	*P*-value	Item-Total Correlation
1	3	1.2	37.5	18.3	20.6	22.4	–	–	–	0.42	0.72
2	3	2.3	55.7	28.4	10.2	3.5	–	–	–	0.20	0.61
3	4	0.3	30.4	32.6	4.3	10.0	22.5	–	–	0.40	0.76
4	4	3.7	42.2	18.9	15.3	10.2	9.7	–	–	0.30	0.78
5	6	0.2	38.0	28.7	13.1	8.9	4.8	3.8	2.5	0.22	0.79

Table I.5. SR Item Classical Statistics—Mathematics Grade 5

Item	Omit %	*P*-value	Item–Total Correlation
1	0.1	0.62	0.54
2	0.8	0.37	0.47
3	0.6	0.65	0.51
4	1.7	0.37	0.62
5	0.2	0.37	0.64
6	1.0	0.48	0.64
7	0.3	0.68	0.55
8	0.1	0.71	0.43
9	0.3	0.32	0.50
10	0.3	0.67	0.58
11	0.2	0.51	0.56
12	1.7	0.64	0.55
13	0.1	0.65	0.39
14	0.1	0.29	0.40
15	0.2	0.48	0.42
16	0.2	0.54	0.51
17	0.5	0.60	0.41
18	3.0	0.46	0.37
19	0.4	0.42	0.54
20	0.1	0.59	0.71
21	2.2	0.41	0.61
22	0.6	0.53	0.61
23	0.0	0.55	0.72
24	0.5	0.30	0.62
25	0.4	0.43	0.60
26	0.2	0.25	0.78
27	1.8	0.25	0.80

Table I.6. CR Item Classical Statistics—Mathematics Grade 5

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	5%	6%	*P*-value	Item-Total Correlation
1	4	0.2	43.6	30.1	13.1	9.5	3.5	–	–	0.25	0.78
2	6	1.8	47.9	5.9	19.1	8.8	6.2	7.6	2.7	0.25	0.80
3	3	5.4	52.4	27.5	11.0	3.8	–	–	–	0.20	0.68
4	3	2.8	54.7	8.4	19.8	14.4	–	–	–	0.30	0.74
5	4	1.5	65.2	8.9	10.4	7.2	6.8	–	–	0.20	0.66

Table I.7. SR Item Classical Statistics—Mathematics Grade 6

Item	Omit %	*P*-value	Item–Total Correlation
1	1.3	0.15	0.51
2	1.2	0.30	0.40
3	0.2	0.59	0.53
4	0.2	0.22	0.50
5	0.3	0.18	0.44
6	1.0	0.11	0.42
7	0.2	0.27	0.39
8	0.7	0.25	0.57
9	2.5	0.26	0.54
10	1.8	0.21	0.45
11	0.8	0.53	0.49
12	0.5	0.32	0.53
13	0.3	0.55	0.54
14	0.1	0.51	0.40
15	0.5	0.62	0.49
16	0.3	0.51	0.48
17	0.2	0.32	0.55
18	2.0	0.43	0.65
19	0.5	0.39	0.72
20	0.2	0.24	0.58
21	0.1	0.41	0.69
22	0.2	0.23	0.33
23	0.4	0.38	0.70
24	0.5	0.38	0.68
25	2.4	0.18	0.79

Table I.8. CR Item Classical Statistics—Mathematics Grade 6

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	5%	6%	*P*-value	Item-Total Correlation
1	3	0.4	18.6	55.6	16.8	8.6	–	–	–	0.38	0.70
2	4	2.4	45.9	25.0	16.0	7.6	3.1	–	–	0.23	0.74
3	6	2.4	60.7	9.3	6.8	7.8	6.4	4.5	2.1	0.18	0.79
4	3	2.0	58.6	18.3	13.2	7.9	–	–	–	0.23	0.76
5	4	1.5	38.6	11.5	16.0	22.2	10.2	–	–	0.38	0.73

Table I.9. SR Item Classical Statistics—Mathematics Grade 7

Item	Omit %	*P*-value	Item–Total Correlation
1	0.7	0.46	0.60
2	0.1	0.51	0.57
3	0.3	0.21	0.58
4	0.5	0.27	0.62
5	0.6	0.32	0.39
6	1.0	0.31	0.66
7	0.4	0.28	0.68
8	0.2	0.41	0.43
9	0.4	0.52	0.45
10	0.1	0.69	0.51
11	0.5	0.45	0.46
12	0.1	0.22	0.39
13	2.9	0.57	0.39
14	0.3	0.30	0.27
15	0.2	0.24	0.47
16	0.2	0.42	0.22
17	0.3	0.57	0.50
18	0.1	0.22	0.25
19	0.5	0.55	0.50
20	0.8	0.14	0.44
21	0.2	0.40	0.48
22	2.2	0.15	0.64
23	0.2	0.29	0.51
24	3.4	0.16	0.75
25	0.2	0.35	0.73
26	1.7	0.29	0.84

Table I.10. CR Item Classical Statistics—Mathematics Grade 7

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	5%	6%	*P*-value	Item-Total Correlation
1	3	3.4	62.0	23.7	8.2	2.6	–	–	–	0.16	0.75
2	6	1.7	31.7	26.9	9.3	9.9	10.0	7.0	3.5	0.29	0.84
3	3	3.0	60.2	21.1	10.6	5.1	–	–	–	0.19	0.65
4	4	2.2	53.5	18.5	7.1	8.7	10.0	–	–	0.25	0.78
5	4	2.6	38.4	14.5	14.3	12.3	17.9	–	–	0.38	0.81

Table I.11. SR Item Classical Statistics—Mathematics Grade 8

Item	Omit %	*P*-value	Item–Total Correlation
1	0.6	0.41	0.61
2	0.9	0.36	0.60
3	0.2	0.43	0.67
4	0.8	0.33	0.33
5	0.6	0.14	0.54
6	0.2	0.35	0.50
7	0.2	0.26	0.49
8	0.1	0.41	0.32
9	0.2	0.35	0.45
10	0.1	0.29	0.44
11	0.9	0.41	0.57
12	0.6	0.27	0.48
13	0.2	0.46	0.52
14	0.2	0.49	0.30
15	0.2	0.78	0.39
16	0.1	0.39	0.27
17	0.3	0.45	0.44
18	1.4	0.27	0.71
19	0.7	0.40	0.68
20	0.2	0.64	0.58
21	0.2	0.46	0.51
22	0.1	0.40	0.52
23	0.4	0.33	0.73
24	0.6	0.24	0.76
25	0.6	0.26	0.81
26	4.7	0.09	0.71

Table I.12. CR Item Classical Statistics—Mathematics Grade 8

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	5%	6%	*P*-value	Item-Total Correlation
1	4	0.6	41.7	33.6	12.7	8.8	2.6	–	–	0.24	0.76
2	4	0.6	57.8	13.9	5.8	8.1	13.8	–	–	0.26	0.81
3	6	4.7	73.1	5.2	10.8	1.5	1.4	1.9	1.5	0.09	0.71
4	3	5.3	84.8	4.1	3.9	1.9	–	–	–	0.06	0.58
5	3	1.8	61.4	16.3	7.5	12.9	–	–	–	0.23	0.70

Table I.13. SR Item Classical Statistics—ELA Grade 3

Item	Omit %	*P*-value	Item–Total Correlation
1	0.6	0.52	0.49
2	0.9	0.36	0.55
3	0.8	0.40	0.55
4	0.7	0.49	0.61
5	0.8	0.57	0.50
6	0.5	0.63	0.61
7	0.6	0.54	0.68
8	0.6	0.57	0.53
9	0.1	0.57	0.59
10	0.2	0.60	0.54
11	0.3	0.56	0.68
12	0.5	0.54	0.61
13	0.3	0.44	0.46
14	0.1	0.72	0.60
15	0.1	0.61	0.61
16	0.1	0.57	0.71
17	0.1	0.61	0.63
18	0.1	0.68	0.52
19	0.2	0.50	0.50

Table I.14. CR Item Classical Statistics—ELA Grade 3

Item	Max. Points	Omit %	0%	1%	2%	3%	*P*-value	Item-Total Correlation
PCR_1_WE	3	2.50	48.68	38.93	9.74	0.16	0.20	0.67
PCR_1_WKL	3	2.50	48.49	43.82	5.04	0.15	0.18	0.59
PCR_2_WE	3	1.15	51.17	43.34	4.23	0.12	0.17	0.72
PCR_2_WKL	3	1.15	45.29	45.67	7.65	0.24	0.21	0.61

Table I.15. SR Item Classical Statistics—ELA Grade 4

Item	Omit %	*P*-value	Item–Total Correlation
1	0.3	0.64	0.54
2	0.5	0.25	0.47
3	0.4	0.38	0.45
4	0.4	0.24	0.37
5	0.1	0.35	0.36
6	0.0	0.41	0.41
7	0.1	0.65	0.62
8	0.1	0.52	0.49
9	0.4	0.40	0.53
10	1.6	0.41	0.46
11	0.6	0.31	0.55
12	0.9	0.40	0.36
13	1.2	0.43	0.68
14	0.7	0.45	0.38
15	1.4	0.38	0.47
16	0.0	0.76	0.55
17	0.0	0.68	0.52
18	0.1	0.39	0.37
19	0.1	0.69	0.63
20	0.1	0.29	0.37
21	0.0	0.49	0.51

Table I.16. CR Item Classical Statistics—ELA Grade 4

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	*P*-value	Item-Total Correlation
PCR_1_WE	3	0.98	29.56	38.11	25.93	5.42	–	0.35	0.75
PCR_1_WKL	3	0.98	43.71	35.21	16.52	3.59	–	0.26	0.70
PCR_2_WE	4	0.68	35.57	51.24	11.05	1.28	0.18	0.19	0.74
PCR_2_WKL	3	0.68	51.18	37.59	9.16	1.39	–	0.20	0.66

Table I.17. SR Item Classical Statistics—ELA Grade 5

Item	Omit %	*P*-value	Item–Total Correlation
1	0.1	0.78	0.54
2	0.1	0.58	0.58
3	0.2	0.49	0.55
4	0.3	0.47	0.51
5	0.2	0.33	0.37
6	0.2	0.62	0.67
7	0.2	0.58	0.66
8	0.0	0.61	0.56
9	0.0	0.34	0.36
10	0.0	0.45	0.51
11	0.0	0.33	0.37
12	0.1	0.43	0.49
13	0.1	0.37	0.43
14	0.2	0.52	0.53
15	0.1	0.57	0.55
16	0.4	0.33	0.56
17	0.2	0.31	0.26
18	0.4	0.70	0.60
19	0.6	0.43	0.48
20	0.9	0.24	0.32

Table I.18. CR Item Classical Statistics—ELA Grade 5

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	*P*-value	Item-Total Correlation
PCR_1_WE	3	0.58	21.44	38.98	34.90	4.11	–	0.40	0.78
PCR_1_WKL	3	0.58	38.48	38.45	19.97	2.52	–	0.29	0.74
PCR_2_WE	4	1.87	34.95	45.44	16.13	1.48	0.14	0.21	0.77
PCR_2_WKL	3	1.87	41.06	41.88	13.55	1.64	–	0.25	0.72

Table I.19. SR Item Classical Statistics—ELA Grade 6

Item	Omit %	*P*-value	Item–Total Correlation
1	0.3	0.43	0.50
2	0.4	0.45	0.49
3	0.4	0.46	0.46
4	0.5	0.49	0.49
5	0.2	0.64	0.45
6	0.2	0.43	0.64
7	0.2	0.45	0.53
8	0.2	0.38	0.50
9	0.2	0.71	0.57
10	0.2	0.58	0.42
11	0.1	0.36	0.39
12	0.1	0.50	0.54
13	0.2	0.51	0.46
14	0.2	0.68	0.64
15	0.1	0.55	0.56
16	0.1	0.41	0.54
17	0.1	0.50	0.58
18	0.1	0.57	0.55
19	0.1	0.33	0.38
20	0.1	0.48	0.38
21	0.0	0.52	0.51
22	0.1	0.65	0.55

Table I.20. CR Item Classical Statistics—ELA Grade 6

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	*P*-value	Item-Total Correlation
PCR_1_WE	4	1.20	27.86	33.14	24.82	10.58	2.40	0.31	0.80
PCR_1_WKL	3	1.20	41.54	29.71	20.26	7.30	–	0.31	0.76
PCR_2_WE	4	1.40	30.19	49.17	15.94	3.23	0.07	0.23	0.78
PCR_2_WKL	3	1.40	37.36	41.20	16.41	3.63	–	0.28	0.76

Table I.21. SR Item Classical Statistics—ELA Grade 7

Item	Omit %	*P*-value	Item–Total Correlation
1	0.4	0.65	0.61
2	0.5	0.40	0.45
3	0.4	0.60	0.62
4	0.4	0.46	0.55
5	0.1	0.46	0.51
6	0.1	0.46	0.47
7	0.2	0.57	0.59
8	0.2	0.54	0.59
9	0.2	0.48	0.40
10	0.3	0.52	0.54
11	0.2	0.53	0.54
12	0.2	0.45	0.57
13	0.2	0.34	0.40
14	0.2	0.43	0.64
15	0.2	0.58	0.42
16	0.2	0.46	0.38
17	0.3	0.37	0.52
18	0.0	0.51	0.53
19	0.1	0.52	0.59
20	0.1	0.45	0.39
21	0.1	0.56	0.61
22	0.1	0.59	0.62
23	0.1	0.42	0.52

Table I.22. CR Item Classical Statistics—ELA Grade 7

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	*P*-value	Item-Total Correlation
PCR_1_WE	4	1.96	27.78	36.11	25.94	7.12	1.09	0.28	0.80
PCR_1_WKL	3	1.96	33.03	32.52	24.69	7.80	–	0.35	0.79
PCR_2_WE	4	1.56	25.55	36.95	28.00	6.87	1.07	0.29	0.83
PCR_2_WKL	3	1.56	36.74	34.15	21.71	5.84	–	0.32	0.79

Table I.23. SR Item Classical Statistics—ELA Grade 8

Item	Omit %	*P*-value	Item–Total Correlation
1	0.2	0.63	0.59
2	0.3	0.50	0.54
3	0.1	0.26	0.27
4	0.1	0.60	0.61
5	0.0	0.73	0.44
6	0.1	0.56	0.51
7	0.3	0.27	0.37
8	0.3	0.45	0.55
9	0.4	0.30	0.40
10	0.4	0.40	0.56
11	0.4	0.44	0.56
12	0.3	0.72	0.41
13	0.2	0.58	0.58
14	0.3	0.35	0.48
15	0.3	0.52	0.55
16	0.2	0.40	0.35
17	0.3	0.61	0.54
18	0.3	0.55	0.59
19	0.4	0.49	0.54
20	0.4	0.44	0.49
21	0.5	0.61	0.58
22	0.5	0.47	0.44
23	0.6	0.24	0.44

Table I.24. CR Item Classical Statistics—ELA Grade 8

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	*P*-value	Item-Total Correlation
PCR_1_WE	4	2.33	36.04	25.20	23.75	11.17	1.51	0.28	0.84
PCR_1_WKL	3	2.33	32.71	28.95	23.73	12.28	–	0.38	0.83
PCR_2_WE	4	2.52	29.51	27.71	26.67	9.92	3.66	0.31	0.83
PCR_2_WKL	3	2.52	31.72	28.44	25.82	11.51	–	0.38	0.82

Table I.25. SR Item Classical Statistics—CSLA Grade 3

Item	Omit %	*P*-value	Item–Total Correlation
1	1.2	0.58	0.56
2	1.7	0.30	0.45
3	1.5	0.27	0.31
4	2.6	0.29	0.57
5	3.4	0.29	0.55
6	4.2	0.49	0.58
7	5.1	0.30	0.52
8	5.2	0.21	0.25
9	4.6	0.40	0.58
10	1.9	0.52	0.57
11	2.2	0.26	0.37
12	3.0	0.31	0.34
13	2.4	0.41	0.50
14	5.3	0.27	0.46
15	5.5	0.38	0.57
16	6.4	0.27	0.39
17	6.1	0.21	0.28
18	6.5	0.28	0.42
19	4.9	0.09	0.37

Table I.26. CR Item Classical Statistics—CSLA Grade 3

Item	Max. Points	Omit %	0%	1%	2%	3%	*P*-value	Item-Total Correlation
PCR_1_WE	3	4.7	47.3	27.8	14.1	6.1	0.25	0.79
PCR_1_WKL	3	4.7	42.0	33.3	14.4	5.6	0.26	0.66
PCR_2_WE	3	6.7	54.3	26.3	10.1	2.6	0.18	0.70
PCR_2_WKL	3	6.7	41.3	33.7	12.5	5.8	0.25	0.61

Table I.27. SR Item Classical Statistics—CSLA Grade 4

Item	Omit %	*P*-value	Item–Total Correlation
1	1.1	0.63	0.56
2	1.2	0.18	0.46
3	0.9	0.52	0.61
4	1.2	0.28	0.53
5	1.4	0.57	0.64
6	0.9	0.52	0.56
7	0.8	0.55	0.63
8	1.9	0.48	0.60
9	1.5	0.54	0.64
10	1.5	0.62	0.51
11	3.1	0.36	0.47
12	4.1	0.20	0.32
13	6.7	0.34	0.57
14	3.8	0.25	0.45
15	3.9	0.35	0.56
16	4.3	0.19	0.16
17	4.4	0.14	0.16
18	4.7	0.43	0.48
19	6.4	0.15	0.33
20	8.8	0.19	0.51
21	4.9	0.23	0.32

Table I.28. CR Item Classical Statistics—CSLA Grade 4

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	*P*-value	Item-Total Correlation
PCR_1_WE	4	4.9	38.5	32.4	15.7	6.7	1.8	0.23	0.83
PCR_1_WKL	3	4.9	40.0	35.5	12.9	6.6	–	0.27	0.69
PCR_2_WE	3	3.8	40.9	19.4	20.7	15.2	–	0.35	0.75
PCR_2_WKL	3	3.8	39.7	37.1	11.0	8.4	–	0.28	0.61

Table I.29. SR Item Classical Statistics—Science Grade 5

Item	Omit %	*P*-value	Item–Total Correlation
1	0.0	0.14	0.31
2	0.4	0.47	0.49
3	0.4	0.52	0.30
4	0.1	0.56	0.45
5	0.1	0.71	0.43
6	0.2	0.46	0.60
7	0.8	0.25	0.33
8	0.9	0.52	0.30
9	0.1	0.53	0.44
10	0.7	0.15	0.33
11	0.1	0.57	0.44
12	1.2	0.28	0.33
13	0.0	0.74	0.49
14	0.7	0.29	0.20
15	0.8	0.35	0.40
16	0.5	0.47	0.42
17	0.0	0.87	0.38
18	0.3	0.36	0.16
19	0.4	0.50	0.25
20	0.2	0.28	0.24
21	0.3	0.59	0.37
22	0.2	0.60	0.52
23	0.0	0.57	0.33
24	0.0	0.71	0.41
25	0.7	0.55	0.42
26	0.6	0.69	0.47
27	0.1	0.54	0.35

Table I.30. CR Item Classical Statistics—Science Grade 5

Item	Max. Points	Omit %	0%	1%	2%	*P*-value	Item-Total Correlation
1	2	1.5	49.7	27.2	21.5	0.35	0.61
2	2	3.5	44.2	42.6	9.8	0.31	0.55
3	2	1.2	40.1	45.3	13.4	0.36	0.57
4	2	1.6	35.2	34.8	28.4	0.46	0.69
5	2	3.3	50.3	22.4	24.0	0.35	0.71
6	2	1.4	39.9	19.3	39.4	0.49	0.66
7	2	2.5	19.0	36.4	42.1	0.60	0.61
8	2	0.8	43.3	36.5	19.4	0.38	0.64
9	2	3.1	37.0	39.7	20.2	0.40	0.67
10	2	5.7	50.7	33.7	10.0	0.27	0.50
11	2	2.6	61.7	32.7	3.0	0.19	0.53
12	2	0.8	49.9	15.5	33.8	0.42	0.67

Table I.31. SR Item Classical Statistics—Science Grade 8

Item	Omit %	*P*-value	Item–Total Correlation
1	0.0	0.53	0.37
2	0.0	0.68	0.46
3	0.2	0.46	0.44
4	0.4	0.10	0.31
5	1.3	0.30	0.38
6	0.0	0.60	0.40
7	0.1	0.11	0.28
8	0.2	0.32	0.51
9	0.2	0.20	0.23
10	0.0	0.73	0.33
11	0.4	0.41	0.45
12	0.1	0.48	0.34
13	0.2	0.36	0.34
14	0.1	0.41	0.41
15	0.1	0.60	0.47
16	0.1	0.34	0.32
17	0.1	0.36	0.17
18	0.1	0.29	0.38
19	1.4	0.36	0.33
20	0.2	0.27	0.38
21	0.1	0.18	0.30
22	0.0	0.34	0.31
23	0.1	0.56	0.39
24	0.2	0.53	0.38
25	0.2	0.39	0.26
26	0.2	0.23	0.35
27	0.2	0.74	0.46
28	0.4	0.54	0.50
29	0.3	0.46	0.52
30	0.1	0.38	0.26
31	0.0	0.50	0.39
32	0.2	0.75	0.48
33	0.0	0.73	0.47

Table I.32. CR Item Classical Statistics—Science Grade 8

Item	Max. Points	Omit %	0%	1%	2%	*P*-value	Item-Total Correlation
1	2	2.1	72.6	17.9	7.4	0.16	0.62
2	2	2.9	80.9	13.6	2.7	0.09	0.50
3	2	3.8	65.7	19.0	11.6	0.21	0.64
4	2	3.1	47.5	26.3	23.1	0.36	0.55
5	2	3.3	61.9	28.9	5.9	0.20	0.56
6	2	2.2	61.7	24.0	12.1	0.24	0.57
7	2	2.7	35.9	17.6	43.8	0.53	0.68
8	2	2.1	48.1	19.0	30.7	0.40	0.69
9	2	2.0	43.3	29.9	24.8	0.40	0.69
10	2	4.3	74.9	9.1	11.7	0.16	0.53
11	2	3.2	79.2	14.2	3.4	0.10	0.48
12	2	4.8	58.6	28.0	8.6	0.23	0.59
13	2	3.7	54.4	30.8	11.1	0.27	0.59
14	2	3.8	58.1	19.3	18.8	0.28	0.69

Table I.33. SR Item Classical Statistics—Science Grade 11

Item	Omit %	*P*-value	Item–Total Correlation
1	1.2	0.30	0.47
2	0.1	0.49	0.48
3	0.5	0.47	0.54
4	0.4	0.43	0.29
5	0.4	0.61	0.56
6	0.1	0.07	0.23
7	0.1	0.65	0.59
8	0.4	0.42	0.44
9	0.6	0.16	0.38
10	0.1	0.48	0.48
11	0.0	0.39	0.45
12	0.2	0.33	0.18
13	0.3	0.54	0.46
14	0.2	0.33	0.36
15	0.3	0.80	0.38
16	0.3	0.43	0.30
17	0.3	0.51	0.39
18	0.0	0.29	0.24
19	0.0	0.57	0.34
20	0.0	0.46	0.43
21	0.1	0.41	0.38
22	0.0	0.60	0.36

Table I.34. CR Item Classical Statistics—Science Grade 11

Item	Max. Points	Omit %	0%	1%	2%	*P*-value	Item-Total Correlation
1	2	9.1	78.4	10.6	1.9	0.07	0.36
2	2	6.3	45.9	21.5	26.3	0.37	0.67
3	2	2.6	36.8	22.4	38.2	0.49	0.68
4	2	3.9	44.4	39.6	12.0	0.32	0.62
5	2	5.2	54.8	27.2	12.8	0.26	0.67
6	2	5.0	32.2	26.8	36.0	0.49	0.69
7	2	2.9	47.4	44.7	5.0	0.27	0.54
8	2	6.3	45.5	33.8	14.4	0.31	0.62
9	2	3.9	83.0	10.8	2.3	0.08	0.38
10	2	6.8	79.9	11.5	1.7	0.07	0.42

Appendix J: Scree Plots

Figure J.1. Scree Plot—Mathematics Grade 3

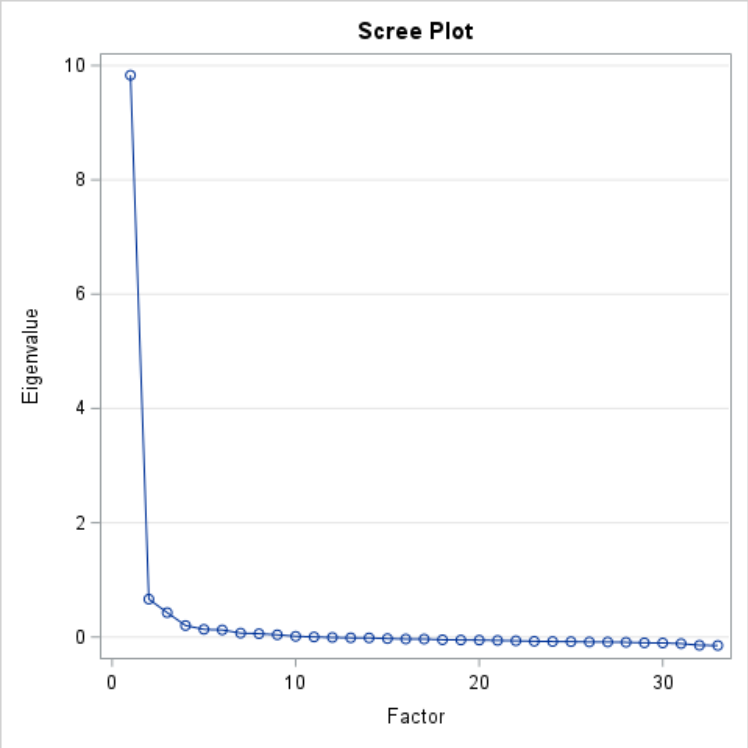


Figure J.2. Scree Plot—Mathematics Grade 4

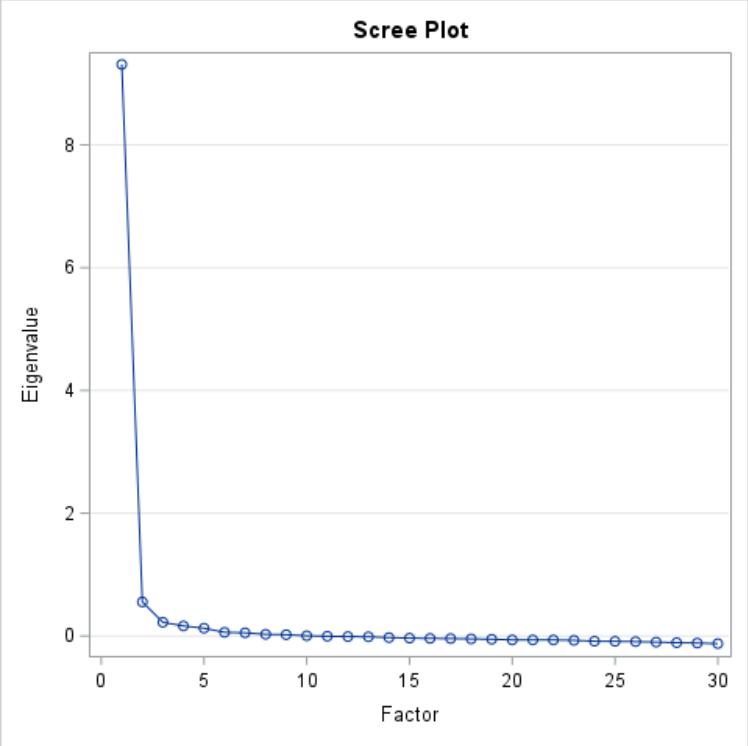


Figure J.3. Scree Plot—Mathematics Grade 5

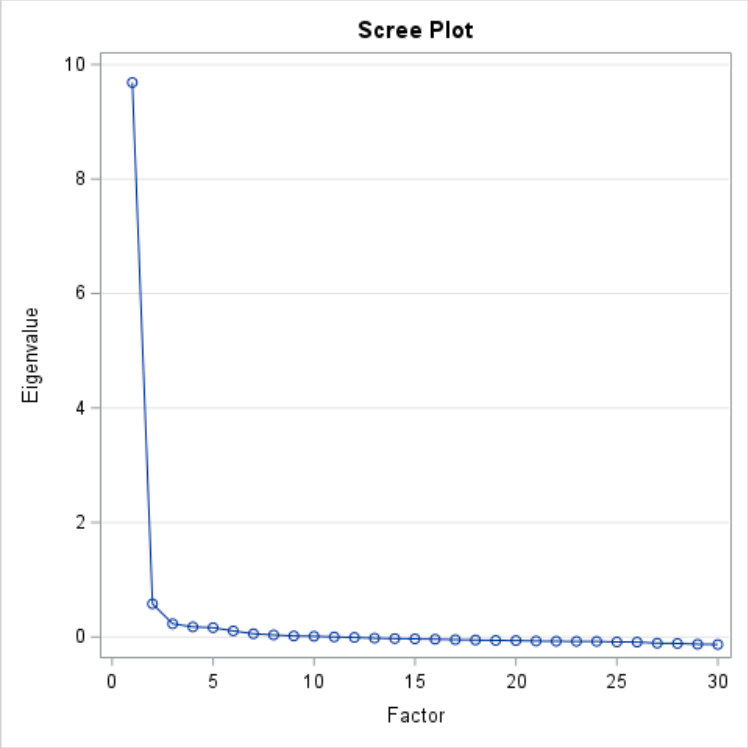


Figure J.4. Scree Plot—Mathematics Grade 6

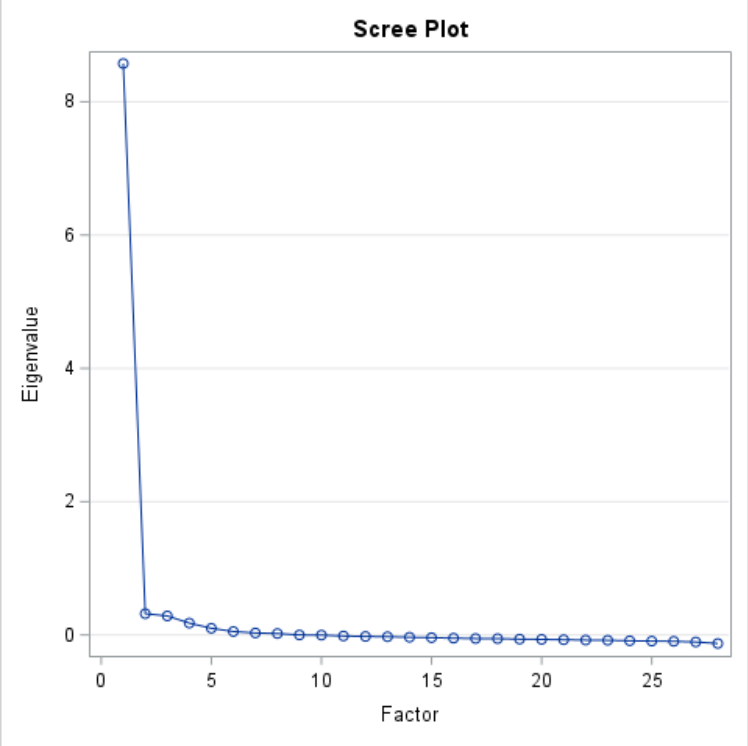


Figure J.5. Scree Plot—Mathematics Grade 7

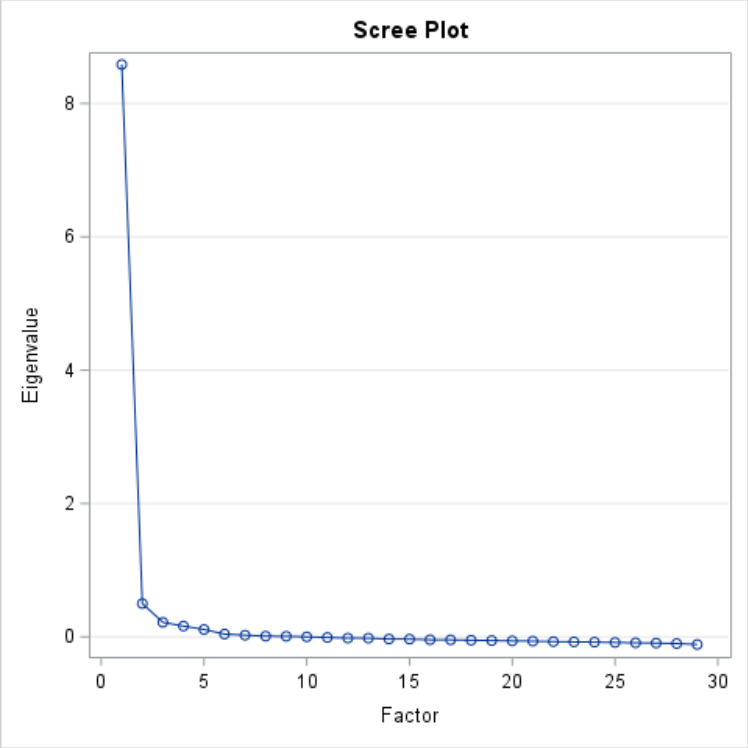


Figure J.6. Scree Plot—Mathematics Grade 8

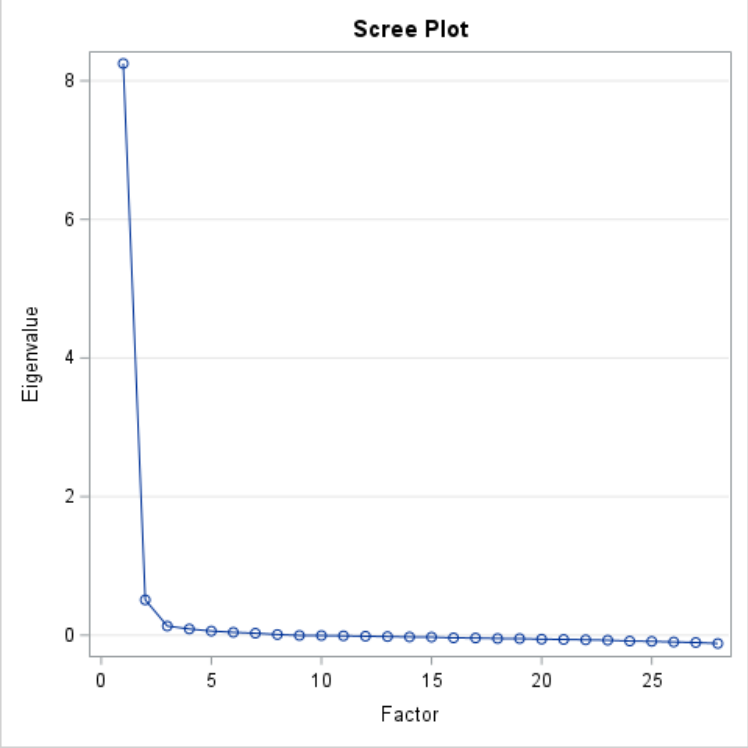


Figure J.7. Scree Plot—ELA Grade 3

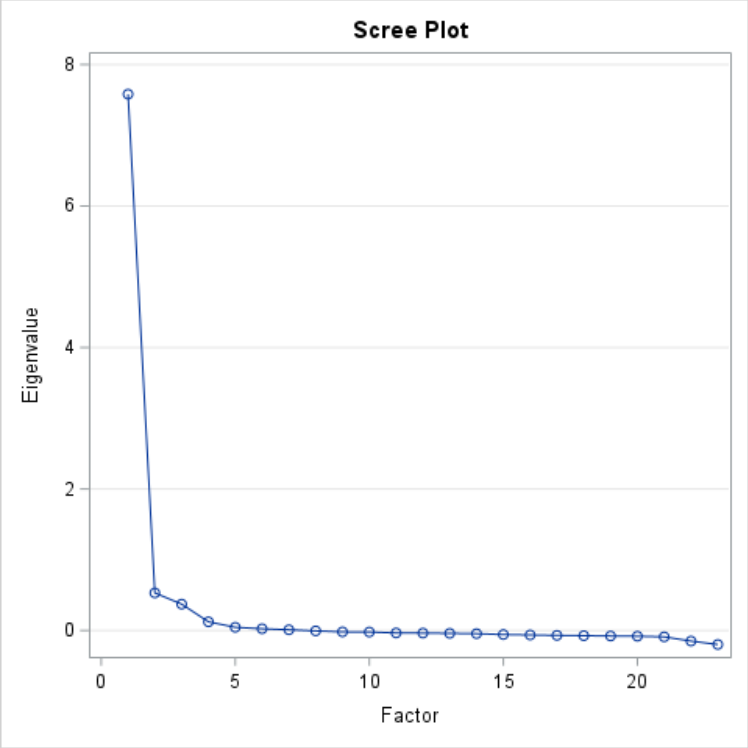


Figure J.8. Scree Plot—ELA Grade 4

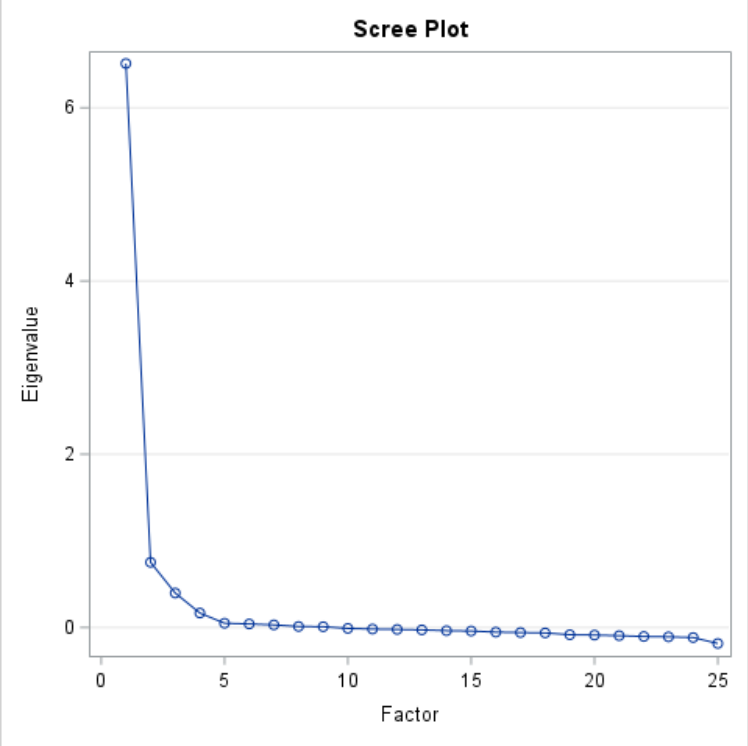


Figure J.9. Scree Plot—ELA Grade 5

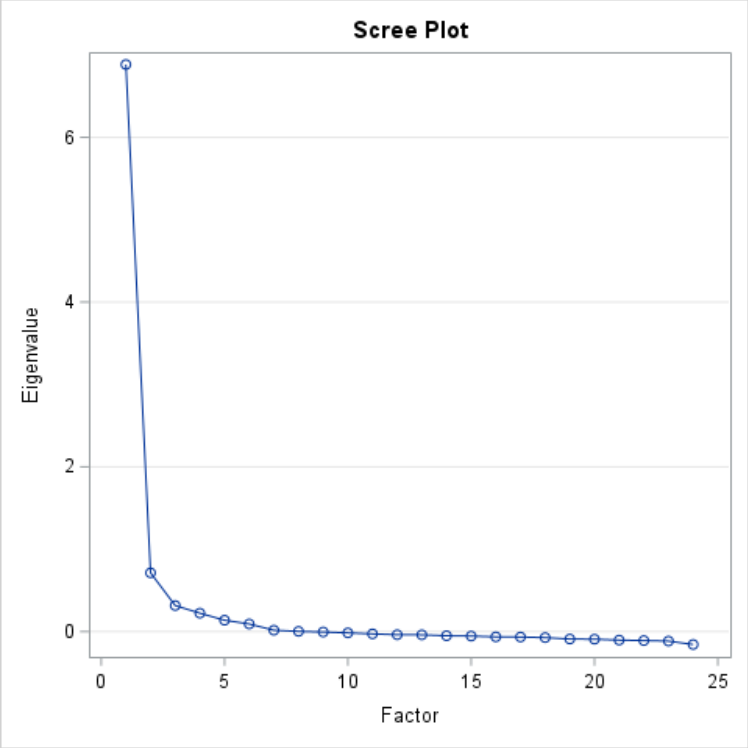


Figure J.10. Scree Plot—ELA Grade 6

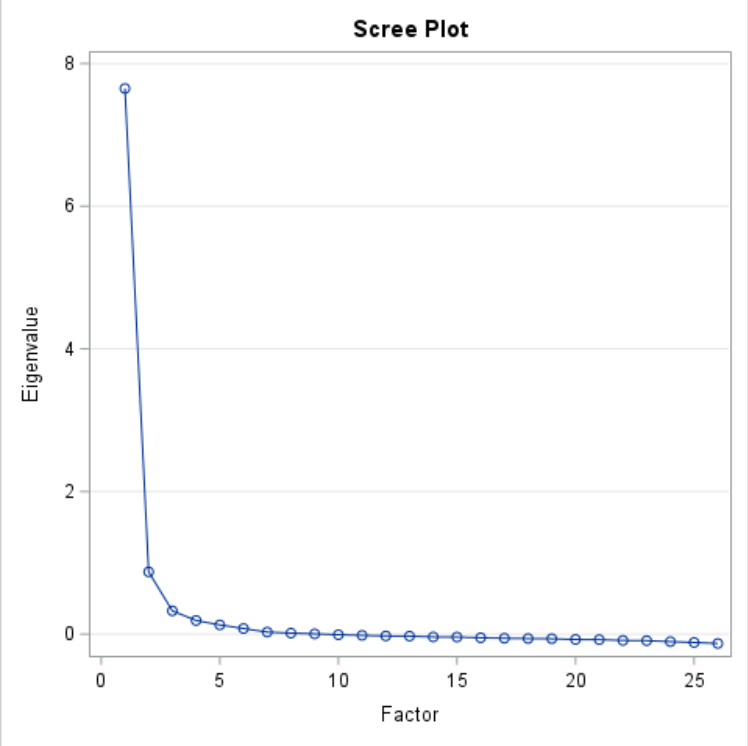


Figure J.11. Scree Plot—ELA Grade 7

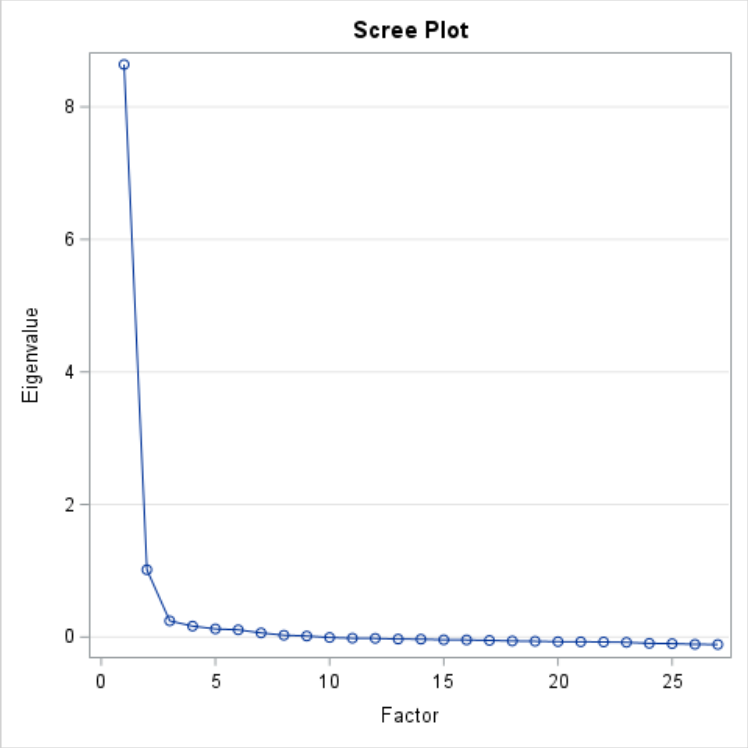


Figure J.12. Scree Plot—ELA Grade 8

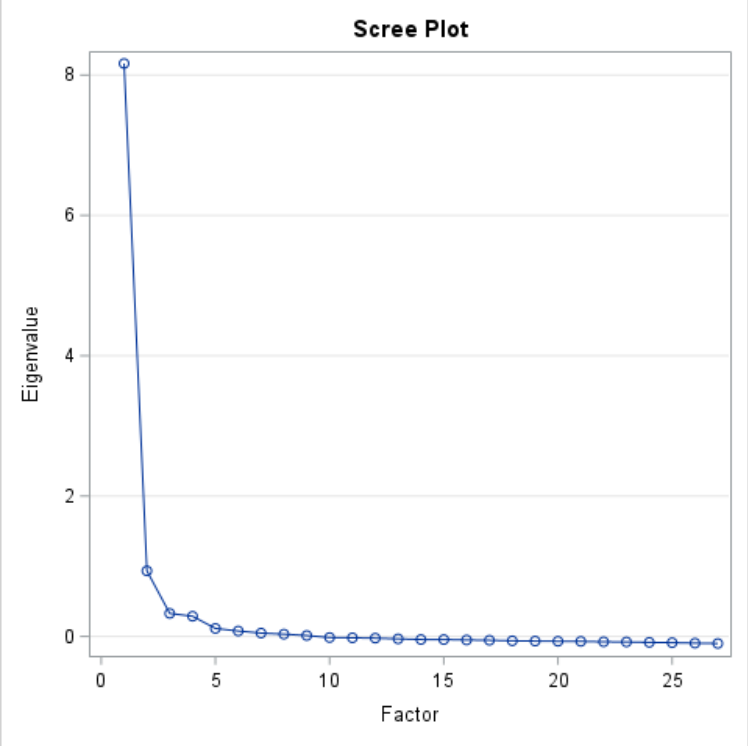


Figure J.13. Scree Plot—CSLA Grade 3

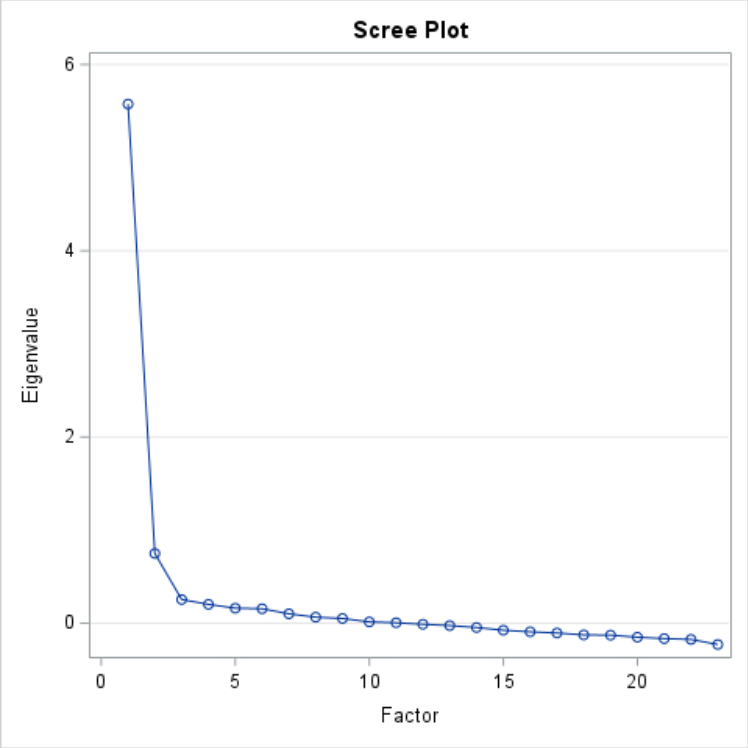


Figure J.14. Scree Plot—CSLA Grade 4

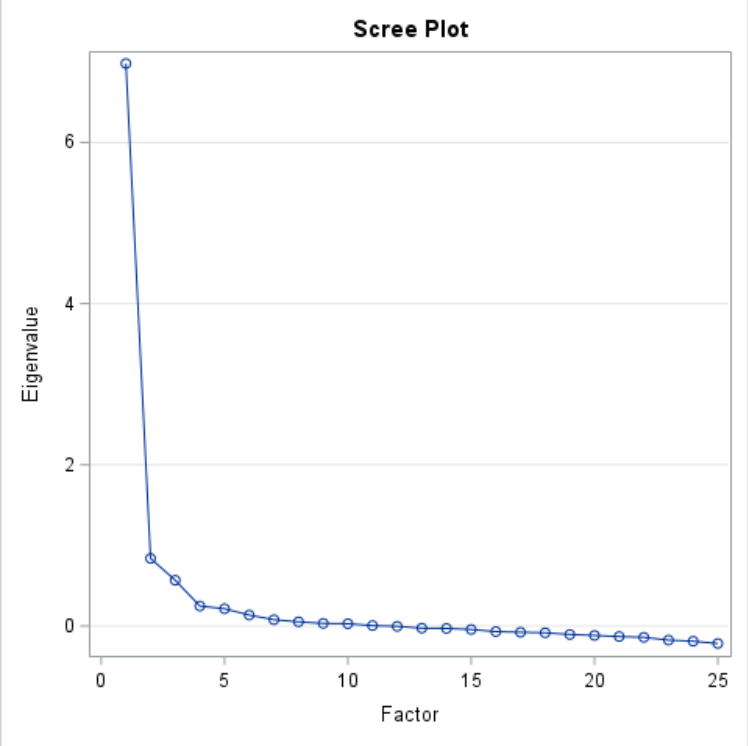


Figure J.15. Scree Plot—Science Grade 5

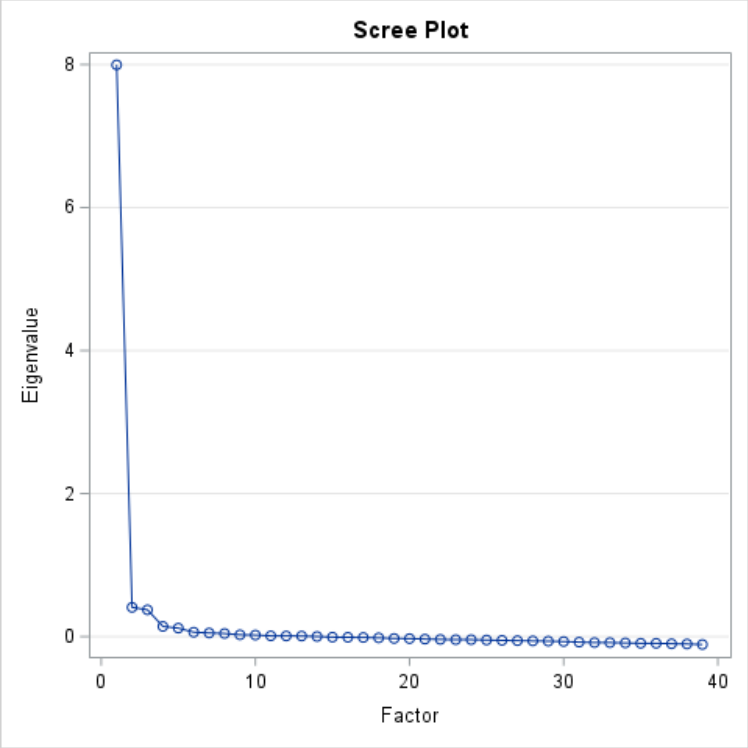


Figure J.16. Scree Plot—Science Grade 8

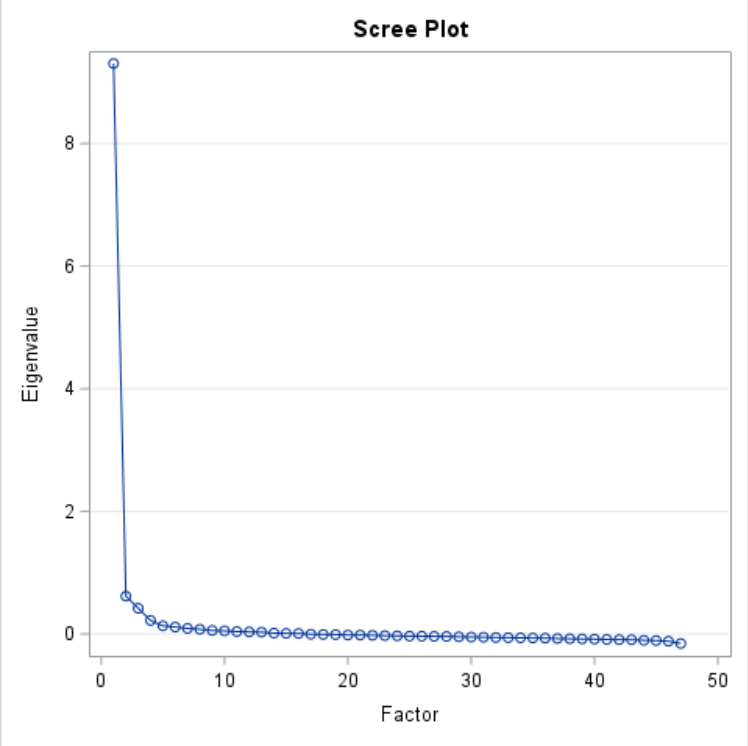
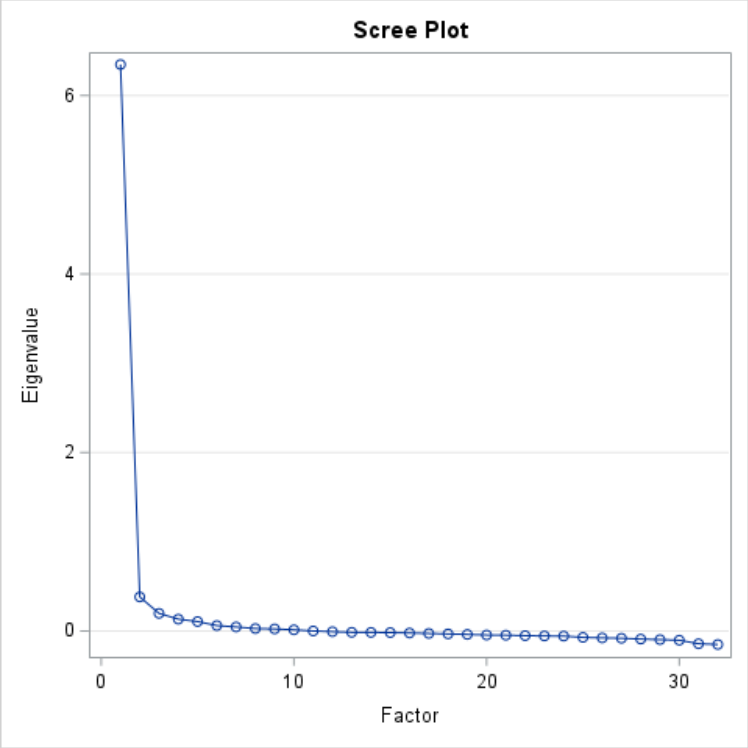


Figure J.17. Scree Plot—Science Grade 11



Appendix K: IRT Item-Level Statistics

Table K.1. Operational Item Parameter Estimates—Mathematics Grade 3

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	D6	D7	Misfit Flag
1	CR	GPC	0.785	0.953	0	-0.220	0.555	-0.334	–	–	–	No
2	CR	GPC	0.455	1.709	0	2.715	0.282	-0.534	-2.463	–	–	No
3	CR	GPC	0.767	0.547	0	0.153	0.117	-0.271	–	–	–	No
4	XI	GPC	0.708	0.955	0	1.543	-1.543	–	–	–	–	No
5	XI	GPC	0.583	-0.149	0	0.738	-0.738	–	–	–	–	No
6	XI	GPC	0.722	0.906	0	0.624	-0.624	–	–	–	–	No
7	XI	GPC	0.428	0.690	0	1.361	-0.803	0.395	-0.953	–	–	No
8	XI	GPC	0.644	0.619	0	2.276	0.798	-0.716	0.402	-2.250	-0.510	No
9	XI	2PL	0.791	-0.079	–	–	–	–	–	–	–	No
10	XI	2PL	0.927	-0.771	–	–	–	–	–	–	–	No
11	XI	2PL	1.124	0.203	–	–	–	–	–	–	–	No
12	XI	2PL	1.095	0.888	–	–	–	–	–	–	–	No
13	XI	2PL	0.689	-0.840	–	–	–	–	–	–	–	No
14	XI	2PL	0.601	1.206	–	–	–	–	–	–	–	No
15	XI	2PL	0.616	1.283	–	–	–	–	–	–	–	No
16	XI	2PL	0.628	-0.824	–	–	–	–	–	–	–	No
17	XI	2PL	0.861	0.719	–	–	–	–	–	–	–	No
18	XI	2PL	1.094	0.161	–	–	–	–	–	–	–	No
19	XI	2PL	1.049	-1.125	–	–	–	–	–	–	–	No
20	XI	2PL	0.702	0.025	–	–	–	–	–	–	–	No
21	XI	2PL	0.696	-1.703	–	–	–	–	–	–	–	No
22	SR	2PL	0.914	-1.309	–	–	–	–	–	–	–	No
23	SR	2PL	0.509	-0.236	–	–	–	–	–	–	–	No
24	SR	2PL	0.820	-0.721	–	–	–	–	–	–	–	No
25	SR	2PL	0.838	-1.325	–	–	–	–	–	–	–	No
26	SR	2PL	0.431	-2.344	–	–	–	–	–	–	–	No
27	SR	2PL	0.629	-0.240	–	–	–	–	–	–	–	No
28	SR	2PL	0.791	-1.646	–	–	–	–	–	–	–	No
29	SR	2PL	0.463	0.032	–	–	–	–	–	–	–	No
30	SR	2PL	0.885	-0.623	–	–	–	–	–	–	–	No
31	SR	2PL	0.586	-1.999	–	–	–	–	–	–	–	No
32	SR	2PL	0.838	-1.324	–	–	–	–	–	–	–	No
33	SR	2PL	0.880	-0.378	–	–	–	–	–	–	–	No

Table K.2. Operational Item Parameter Estimates—Mathematics Grade 4

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	D6	D7	Misfit Flag
1	XI	GPC	0.674	-1.252	0	0.403	-0.403	–	–	–	–	No
2	XI	GPC	0.404	-0.335	0	0.751	-0.751	–	–	–	–	No
3	XI	GPC	0.444	0.083	0	0.923	-2.323	1.014	0.386	–	–	No
4	XI	GPC	0.594	0.214	0	0.093	0.419	-0.512	–	–	–	No
5	XI	GPC	0.801	-0.188	0	0.486	-0.486	–	–	–	–	No
6	XI	GPC	0.870	-0.275	0	1.106	-1.106	–	–	–	–	No
7	XI	GPC	0.636	0.443	0	0.767	-0.767	–	–	–	–	No
8	XI	GPC	1.043	0.916	0	0.317	-0.317	–	–	–	–	No
9	XI	GPC	0.577	1.012	0	0.800	-0.237	-0.563	–	–	–	No
10	XI	GPC	0.733	0.550	0	0.385	0.442	-0.209	-0.618	–	–	No
11	XI	GPC	0.653	0.966	0	1.375	0.420	0.005	-0.459	-0.137	-1.204	No
12	XI	2PL	0.970	0.978	–	–	–	–	–	–	–	No
13	XI	2PL	1.111	0.853	–	–	–	–	–	–	–	No
14	XI	2PL	1.317	1.064	–	–	–	–	–	–	–	No
15	XI	2PL	0.533	-0.769	–	–	–	–	–	–	–	No
16	XI	2PL	0.956	0.381	–	–	–	–	–	–	–	No
17	XI	2PL	1.040	0.385	–	–	–	–	–	–	–	No
18	XI	2PL	0.593	-0.013	–	–	–	–	–	–	–	No
19	XI	2PL	0.344	-1.932	–	–	–	–	–	–	–	No
20	SR	2PL	0.549	-0.597	–	–	–	–	–	–	–	No
21	SR	2PL	0.575	0.294	–	–	–	–	–	–	–	No
22	SR	2PL	0.779	-0.900	–	–	–	–	–	–	–	No
23	SR	2PL	0.688	-0.718	–	–	–	–	–	–	–	No
24	SR	2PL	0.723	-1.429	–	–	–	–	–	–	–	No
25	SR	2PL	0.693	0.123	–	–	–	–	–	–	–	No
26	SR	2PL	0.521	0.123	–	–	–	–	–	–	–	No
27	SR	2PL	0.903	-0.155	–	–	–	–	–	–	–	No
28	SR	2PL	0.880	-0.814	–	–	–	–	–	–	–	No
29	SR	2PL	1.121	-1.179	–	–	–	–	–	–	–	No
30	SR	2PL	0.504	1.278	–	–	–	–	–	–	–	No

Table K.3. Operational Item Parameter Estimates—Mathematics Grade 5

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	D6	D7	Misfit Flag
1	CR	GPC	0.723	1.373	0	0.934	-0.034	-0.900	–	–	–	No
2	CR	GPC	0.611	0.704	0	-0.473	0.666	-0.194	–	–	–	No
3	CR	GPC	0.383	1.748	0	-0.941	1.092	0.282	-0.432	–	–	No
4	XI	GPC	0.851	-0.385	0	0.344	-0.344	–	–	–	–	No
5	XI	GPC	0.580	0.365	0	0.493	-0.493	–	–	–	–	No
6	XI	GPC	0.522	1.362	0	-0.624	2.848	-1.030	-0.368	0.034	-0.860	No
7	XI	GPC	0.803	1.135	0	1.249	0.101	-0.197	-1.153	–	–	No
8	XI	GPC	0.676	-0.355	0	0.451	-0.451	–	–	–	–	No
9	XI	GPC	0.876	-0.501	0	0.315	-0.315	–	–	–	–	No
10	XI	GPC	0.640	0.842	0	0.122	-0.122	–	–	–	–	No
11	XI	GPC	0.476	0.442	0	0.455	-0.455	–	–	–	–	No
12	XI	2PL	0.796	-0.304	–	–	–	–	–	–	–	No
13	XI	2PL	0.501	1.021	–	–	–	–	–	–	–	No
14	XI	2PL	0.676	-1.277	–	–	–	–	–	–	–	No
15	XI	2PL	0.876	0.233	–	–	–	–	–	–	–	No
16	XI	2PL	1.064	0.261	–	–	–	–	–	–	–	No
17	XI	2PL	0.925	-0.211	–	–	–	–	–	–	–	No
18	XI	2PL	0.867	-0.634	–	–	–	–	–	–	–	No
19	XI	2PL	0.533	-1.614	–	–	–	–	–	–	–	No
20	XI	2PL	0.616	0.767	–	–	–	–	–	–	–	No
21	SR	2PL	1.049	-0.923	–	–	–	–	–	–	–	No
22	SR	2PL	0.844	-0.213	–	–	–	–	–	–	–	No
23	SR	2PL	0.849	-0.812	–	–	–	–	–	–	–	No
24	SR	2PL	0.524	-1.030	–	–	–	–	–	–	–	No
25	SR	2PL	0.488	1.259	–	–	–	–	–	–	–	No
26	SR	2PL	0.463	-0.045	–	–	–	–	–	–	–	No
27	SR	2PL	0.642	-0.344	–	–	–	–	–	–	–	No
28	SR	2PL	0.513	-1.134	–	–	–	–	–	–	–	No
29	SR	2PL	0.538	0.037	–	–	–	–	–	–	–	No
30	SR	2PL	0.659	0.150	–	–	–	–	–	–	–	No

Table K.4. Operational Item Parameter Estimates—Mathematics Grade 6

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	D6	D7	Misfit Flag
1	CR	GPC	0.826	0.890	0	0.365	0.135	-0.501	–	–	–	No
2	CR	GPC	0.437	0.144	0	0.087	0.688	0.957	-1.732	–	–	No
3	XI	GPC	0.889	0.194	0	0.247	-0.247	–	–	–	–	No
4	XI	GPC	0.598	1.190	0	0.021	0.752	0.942	0.151	-0.638	-1.228	No
5	XI	GPC	0.772	0.189	0	2.033	-0.714	-1.319	–	–	–	No
6	XI	GPC	0.638	1.258	0	0.784	-0.784	–	–	–	–	No
7	XI	GPC	0.790	0.146	0	0.345	-0.345	–	–	–	–	No
8	XI	GPC	0.275	2.550	0	1.883	-1.883	–	–	–	–	No
9	XI	GPC	0.478	0.399	0	1.487	0.225	-0.720	-0.991	–	–	No
10	XI	GPC	0.802	0.960	0	1.238	0.298	-0.239	-1.296	–	–	No
11	XI	2PL	0.838	1.039	–	–	–	–	–	–	–	No
12	XI	2PL	0.483	1.132	–	–	–	–	–	–	–	No
13	XI	2PL	0.839	-0.746	–	–	–	–	–	–	–	No
14	XI	2PL	0.704	1.295	–	–	–	–	–	–	–	No
15	XI	2PL	0.649	1.564	–	–	–	–	–	–	–	No
16	XI	2PL	0.762	2.042	–	–	–	–	–	–	–	No
17	XI	2PL	0.804	0.312	–	–	–	–	–	–	–	No
18	XI	2PL	0.501	1.098	–	–	–	–	–	–	–	No
19	XI	2PL	0.884	0.733	–	–	–	–	–	–	–	No
20	XI	2PL	0.820	0.723	–	–	–	–	–	–	–	No
21	XI	2PL	0.677	1.230	–	–	–	–	–	–	–	No
22	SR	2PL	0.664	-0.350	–	–	–	–	–	–	–	No
23	SR	2PL	0.705	0.669	–	–	–	–	–	–	–	No
24	SR	2PL	0.838	-0.200	–	–	–	–	–	–	–	No
25	SR	2PL	0.501	-0.465	–	–	–	–	–	–	–	No
26	SR	2PL	0.747	-0.812	–	–	–	–	–	–	–	No
27	SR	2PL	0.589	0.065	–	–	–	–	–	–	–	No
28	SR	2PL	0.863	0.511	–	–	–	–	–	–	–	No
29	SR	2PL	1.147	0.000	–	–	–	–	–	–	–	No

Table K.5. Operational Item Parameter Estimates—Mathematics Grade 7

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	D6	D7	Misfit Flag
1	CR	GPC	0.575	0.889	0	0.216	-0.505	0.244	0.046	–	–	No
2	CR	GPC	0.629	0.013	0	0.010	0.443	-0.205	-0.248	–	–	No
3	CR	GPC	0.485	1.426	0	0.103	0.597	-0.701	–	–	–	No
4	XI	GPC	0.366	1.963	0	-1.175	1.175	–	–	–	–	No
5	XI	GPC	0.372	0.681	0	0.657	-0.657	–	–	–	–	No
6	XI	GPC	0.598	0.804	0	1.420	-0.054	0.491	0.017	-0.703	-1.171	No
7	XI	GPC	1.012	1.598	0	0.487	-0.487	–	–	–	–	No
8	XI	GPC	0.482	1.265	0	0.848	-0.848	–	–	–	–	No
9	XI	GPC	0.544	0.790	0	1.297	0.213	-0.249	-1.261	–	–	No
10	XI	GPC	1.094	1.322	0	0.906	-0.002	-0.904	–	–	–	No
11	XI	2PL	1.031	-0.114	–	–	–	–	–	–	–	No
12	XI	2PL	0.842	-0.047	–	–	–	–	–	–	–	No
13	XI	2PL	0.964	1.280	–	–	–	–	–	–	–	No
14	XI	2PL	1.043	0.853	–	–	–	–	–	–	–	No
15	XI	2PL	0.534	0.838	–	–	–	–	–	–	–	No
16	XI	2PL	1.138	0.801	–	–	–	–	–	–	–	No
17	XI	2PL	1.450	0.869	–	–	–	–	–	–	–	No
18	SR	2PL	0.487	0.294	–	–	–	–	–	–	–	No
19	SR	2PL	0.458	-0.075	–	–	–	–	–	–	–	No
20	SR	2PL	1.041	-0.780	–	–	–	–	–	–	–	No
21	SR	2PL	0.558	0.121	–	–	–	–	–	–	–	No
22	SR	2PL	0.555	1.625	–	–	–	–	–	–	–	No
23	SR	2PL	0.454	-0.962	–	–	–	–	–	–	–	No
24	SR	2PL	0.279	1.924	–	–	–	–	–	–	–	No
25	SR	2PL	0.654	1.069	–	–	–	–	–	–	–	No
26	SR	2PL	0.252	0.674	–	–	–	–	–	–	–	No
27	SR	2PL	0.691	-0.500	–	–	–	–	–	–	–	No
28	SR	2PL	0.366	1.908	–	–	–	–	–	–	–	No
29	SR	2PL	0.669	-0.652	–	–	–	–	–	–	–	No

Table K.6. Operational Item Parameter Estimates—Mathematics Grade 8

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	D6	D7	Misfit Flag
1	CR	GPC	0.869	2.202	0	-0.302	0.732	-0.430	–	–	–	No
2	CR	GPC	0.537	1.435	0	-0.119	-0.641	0.759	–	–	–	No
3	XI	GPC	0.818	1.503	0	0.809	-0.809	–	–	–	–	No
4	XI	GPC	0.397	0.754	0	1.167	-0.541	-0.149	-0.477	–	–	No
5	XI	GPC	0.576	2.107	0	-1.074	2.149	-1.620	0.919	0.226	-0.600	No
6	XI	GPC	0.717	1.605	0	1.717	0.057	-0.243	-1.531	–	–	No
7	XI	GPC	0.559	0.837	0	0.362	-0.362	–	–	–	–	No
8	XI	GPC	0.644	-0.545	0	0.593	-0.593	–	–	–	–	No
9	XI	GPC	0.343	0.474	0	0.955	-0.955	–	–	–	–	No
10	XI	GPC	0.431	1.001	0	1.185	-1.185	–	–	–	–	No
11	XI	GPC	0.683	0.908	0	-0.141	-0.012	0.354	-0.201	–	–	No
12	XI	2PL	0.847	0.432	–	–	–	–	–	–	–	No
13	XI	2PL	0.926	0.522	–	–	–	–	–	–	–	No
14	XI	2PL	1.103	0.222	–	–	–	–	–	–	–	No
15	XI	2PL	0.378	1.199	–	–	–	–	–	–	–	No
16	XI	2PL	1.115	1.627	–	–	–	–	–	–	–	No
17	XI	2PL	0.555	0.844	–	–	–	–	–	–	–	No
18	XI	2PL	0.523	1.803	–	–	–	–	–	–	–	No
19	SR	2PL	0.321	0.916	–	–	–	–	–	–	–	No
20	SR	2PL	0.529	1.042	–	–	–	–	–	–	–	No
21	SR	2PL	0.495	1.344	–	–	–	–	–	–	–	No
22	SR	2PL	0.669	0.521	–	–	–	–	–	–	–	No
23	SR	2PL	0.518	1.671	–	–	–	–	–	–	–	No
24	SR	2PL	0.604	0.205	–	–	–	–	–	–	–	No
25	SR	2PL	0.314	-0.184	–	–	–	–	–	–	–	No
26	SR	2PL	0.699	-1.276	–	–	–	–	–	–	–	No
27	SR	2PL	0.289	1.873	–	–	–	–	–	–	–	No
28	SR	2PL	0.495	0.392	–	–	–	–	–	–	–	No

Table K.7. Operational Item Parameter Estimates—ELA Grade 3

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
1	CR	GPC	0.679	2.623	0	2.495	-0.296	-2.199	–	No
2	CR	GPC	0.693	2.483	0	2.296	0.356	-2.652	–	No
3	CR	GPC	0.677	2.377	0	2.419	-0.124	-2.295	–	No
4	CR	GPC	0.971	2.341	0	2.166	-0.261	-1.905	–	No
5	XI	GPC	0.348	-0.083	0	-0.202	0.202	–	–	Yes
6	XI	GPC	0.407	0.682	0	-2.393	2.393	–	–	Yes
7	XI	GPC	0.410	0.553	0	-1.288	1.288	–	–	Yes
8	XI	GPC	0.484	0.112	0	-1.028	1.028	–	–	Yes
9	XI	GPC	0.355	-0.388	0	-0.176	0.176	–	–	Yes
10	XI	GPC	0.485	-0.503	0	-1.271	1.271	–	–	Yes
11	XI	GPC	0.623	-0.088	0	-0.610	0.610	–	–	Yes
12	XI	GPC	0.571	-0.450	0	1.469	-1.469	–	–	No
13	XI	GPC	0.448	-0.258	0	-0.990	0.990	–	–	Yes
14	XI	GPC	0.383	-0.443	0	-1.181	1.181	–	–	Yes
15	XI	GPC	0.701	-0.162	0	-0.025	0.025	–	–	No
16	XI	GPC	0.575	-0.106	0	0.411	-0.411	–	–	No
17	XI	GPC	0.382	0.457	0	1.188	-1.188	–	–	Yes
18	XI	GPC	0.624	-1.006	0	0.285	-0.285	–	–	Yes
19	XI	GPC	0.528	-0.460	0	-0.246	0.246	–	–	No
20	XI	GPC	0.785	-0.197	0	-0.025	0.025	–	–	No
21	XI	GPC	0.527	-0.422	0	-0.808	0.808	–	–	Yes
22	XI	GPC	0.476	-1.013	0	0.842	-0.842	–	–	No
23	XI	GPC	0.520	0.035	0	1.779	-1.779	–	–	Yes

Table K.8. Operational Item Parameter Estimates—ELA Grade 4

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
1	CR	GPC	0.630	1.356	0	1.357	0.039	-1.396	–	No
2	CR	GPC	0.627	0.881	0	1.553	0.146	-1.699	–	No
3	CR	GPC	0.665	1.867	0	1.638	-0.223	-1.415	–	No
4	CR	GPC	0.764	2.110	0	2.645	0.203	-1.137	-1.711	No
5	XI	GPC	0.442	-0.620	0	-1.107	1.107	–	–	Yes
6	XI	GPC	0.456	1.416	0	0.257	-0.257	–	–	Yes
7	XI	GPC	0.340	0.746	0	-0.164	0.164	–	–	Yes
8	XI	GPC	0.286	1.958	0	-0.170	0.170	–	–	Yes
9	XI	GPC	0.207	1.171	0	-3.504	3.504	–	–	Yes
10	XI	GPC	0.303	0.657	0	0.508	-0.508	–	–	No
11	XI	GPC	0.732	-0.642	0	0.340	-0.340	–	–	Yes
12	XI	GPC	0.419	-0.133	0	0.677	-0.677	–	–	No
13	XI	GPC	0.398	0.478	0	-1.697	1.697	–	–	Yes
14	XI	GPC	0.323	0.496	0	-1.466	1.466	–	–	Yes
15	XI	GPC	0.557	0.925	0	0.158	-0.158	–	–	Yes
16	XI	GPC	0.211	0.781	0	-2.381	2.381	–	–	Yes
17	XI	GPC	0.893	0.282	0	0.446	-0.446	–	–	Yes
18	XI	GPC	0.253	0.402	0	-0.247	0.247	–	–	Yes
19	XI	GPC	0.335	0.668	0	-1.259	1.259	–	–	Yes
20	XI	GPC	0.612	-1.114	0	-0.535	0.535	–	–	Yes
21	XI	GPC	0.547	-0.952	0	0.766	-0.766	–	–	Yes
22	XI	GPC	0.249	0.879	0	-0.024	0.024	–	–	Yes
23	XI	GPC	0.712	-0.686	0	-0.876	0.876	–	–	Yes
24	XI	GPC	0.263	1.576	0	-0.547	0.547	–	–	Yes
25	XI	GPC	0.398	0.047	0	-0.452	0.452	–	–	Yes

Table K.9. Operational Item Parameter Estimates—ELA Grade 5

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
1	CR	GPC	0.875	1.411	0	1.452	0.123	-1.575	–	No
2	CR	GPC	0.880	0.900	0	1.672	0.303	-1.975	–	No
3	CR	GPC	0.800	1.719	0	1.662	-0.065	-1.597	–	No
4	CR	GPC	0.843	2.223	0	2.449	0.597	-1.184	-1.862	No
5	XI	GPC	0.622	-0.854	0	-1.244	1.244	–	–	Yes
6	XI	GPC	0.526	-0.083	0	-0.417	0.417	–	–	Yes
7	XI	GPC	0.453	0.310	0	-0.784	0.784	–	–	Yes
8	XI	GPC	0.421	0.404	0	-0.100	0.100	–	–	No
9	XI	GPC	0.259	1.481	0	-0.489	0.489	–	–	No
10	XI	GPC	0.853	-0.197	0	-0.006	0.006	–	–	No
11	XI	GPC	0.729	-0.024	0	-0.446	0.446	–	–	No
12	XI	GPC	0.522	-0.234	0	-0.051	0.051	–	–	No
13	XI	GPC	0.271	1.493	0	0.340	-0.340	–	–	No
14	XI	GPC	0.520	0.565	0	1.059	-1.059	–	–	No
15	XI	GPC	0.281	1.547	0	0.328	-0.328	–	–	Yes
16	XI	GPC	0.389	0.666	0	-0.103	0.103	–	–	No
17	XI	GPC	0.293	0.990	0	-2.183	2.183	–	–	Yes
18	XI	GPC	0.385	0.157	0	-2.402	2.402	–	–	Yes
19	XI	GPC	0.458	-0.051	0	-0.774	0.774	–	–	No
20	XI	GPC	0.534	1.001	0	-0.264	0.264	–	–	No
21	XI	GPC	0.168	2.370	0	-0.652	0.652	–	–	No
22	XI	GPC	0.625	-0.530	0	-0.796	0.796	–	–	Yes
23	XI	GPC	0.332	0.605	0	-2.131	2.131	–	–	Yes
24	XI	GPC	0.286	2.700	0	1.161	-1.161	–	–	No

Table K.10. Operational Item Parameter Estimates—ELA Grade 6

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
1	CR	GPC	0.711	1.092	0	0.973	0.112	-1.086	–	No
2	CR	GPC	0.694	1.215	0	1.716	0.585	-0.546	-1.755	No
3	CR	GPC	0.827	1.266	0	1.491	-0.154	-1.338	–	No
4	CR	GPC	0.868	2.112	0	2.649	0.575	-0.535	-2.689	No
5	XI	GPC	0.398	0.546	0	-0.096	0.096	–	–	Yes
6	XI	GPC	0.367	0.421	0	-0.679	0.679	–	–	No
7	XI	GPC	0.359	0.399	0	0.278	-0.278	–	–	Yes
8	XI	GPC	0.490	0.223	0	1.187	-1.187	–	–	No
9	XI	GPC	0.322	-0.614	0	-2.123	2.123	–	–	Yes
10	XI	GPC	0.657	0.434	0	-0.084	0.084	–	–	No
11	XI	GPC	0.405	0.405	0	-1.034	1.034	–	–	Yes
12	XI	GPC	0.382	0.755	0	-0.993	0.993	–	–	Yes
13	XI	GPC	0.694	-0.827	0	0.624	-0.624	–	–	Yes
14	XI	GPC	0.351	-0.405	0	1.222	-1.222	–	–	Yes
15	XI	GPC	0.264	1.087	0	-1.570	1.570	–	–	Yes
16	XI	GPC	0.397	0.157	0	-2.296	2.296	–	–	Yes
17	XI	GPC	0.469	0.124	0	1.532	-1.532	–	–	Yes
18	XI	GPC	0.839	-0.577	0	0.420	-0.420	–	–	Yes
19	XI	GPC	0.448	-0.046	0	-1.081	1.081	–	–	Yes
20	XI	GPC	0.500	0.619	0	0.306	-0.306	–	–	No
21	XI	GPC	0.482	0.167	0	-0.906	0.906	–	–	No
22	XI	GPC	0.437	-0.134	0	-1.191	1.191	–	–	Yes
23	XI	GPC	0.352	1.591	0	1.433	-1.433	–	–	No
24	XI	GPC	0.353	0.310	0	2.128	-2.128	–	–	No
25	XI	GPC	0.508	0.028	0	1.187	-1.187	–	–	No
26	XI	GPC	0.471	-0.496	0	-1.214	1.214	–	–	No

Table K.11. Operational Item Parameter Estimates—ELA Grade 7

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
1	CR	GPC	0.761	0.821	0	1.129	0.143	-1.272	–	No
2	CR	GPC	0.727	1.407	0	2.024	0.731	-0.809	-1.946	No
3	CR	GPC	0.804	1.004	0	1.207	0.094	-1.301	–	No
4	CR	GPC	0.809	1.314	0	2.037	0.724	-0.891	-1.870	No
5	XI	GPC	0.621	-0.500	0	-1.047	1.047	–	–	Yes
6	XI	GPC	0.349	0.672	0	-0.405	0.405	–	–	No
7	XI	GPC	0.595	-0.283	0	-1.127	1.127	–	–	Yes
8	XI	GPC	0.674	0.307	0	1.286	-1.286	–	–	No
9	XI	GPC	0.372	0.283	0	-2.106	2.106	–	–	Yes
10	XI	GPC	0.357	0.305	0	-0.509	0.509	–	–	Yes
11	XI	GPC	0.545	-0.192	0	-0.651	0.651	–	–	Yes
12	XI	GPC	0.521	-0.049	0	-1.192	1.192	–	–	Yes
13	XI	GPC	0.388	0.266	0	1.992	-1.992	–	–	Yes
14	XI	GPC	0.642	-0.009	0	1.269	-1.269	–	–	No
15	XI	GPC	0.476	-0.037	0	0.206	-0.206	–	–	No
16	XI	GPC	0.444	0.320	0	-2.081	2.081	–	–	Yes
17	XI	GPC	0.290	1.130	0	-0.787	0.787	–	–	No
18	XI	GPC	0.837	0.415	0	0.828	-0.828	–	–	Yes
19	XI	GPC	0.302	-0.380	0	-0.928	0.928	–	–	No
20	XI	GPC	0.249	0.353	0	-0.877	0.877	–	–	No
21	XI	GPC	0.448	0.768	0	0.027	-0.027	–	–	Yes
22	XI	GPC	0.427	0.057	0	-0.794	0.794	–	–	Yes
23	XI	GPC	0.543	0.026	0	-0.448	0.448	–	–	Yes
24	XI	GPC	0.438	0.575	0	2.133	-2.133	–	–	Yes
25	XI	GPC	0.673	-0.170	0	0.451	-0.451	–	–	Yes
26	XI	GPC	0.745	-0.322	0	0.661	-0.661	–	–	Yes
27	XI	GPC	0.450	0.510	0	-0.075	0.075	–	–	Yes

Table K.12. Operational Item Parameter Estimates—ELA Grade 8

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
1	CR	GPC	0.773	0.459	0	0.938	0.053	-0.991	–	No
2	CR	GPC	0.732	1.188	0	1.406	0.890	-0.343	-1.954	No
3	CR	GPC	0.776	0.463	0	0.961	0.150	-1.111	–	No
4	CR	GPC	0.721	0.853	0	1.428	0.689	-0.800	-1.317	No
5	XI	GPC	0.523	-0.628	0	-1.378	1.378	–	–	Yes
6	XI	GPC	0.433	-0.088	0	-0.398	0.398	–	–	No
7	XI	GPC	0.206	2.679	0	1.270	-1.270	–	–	Yes
8	XI	GPC	0.574	-0.511	0	-0.571	0.571	–	–	No
9	XI	GPC	0.368	-1.390	0	-0.551	0.551	–	–	No
10	XI	GPC	0.486	-0.450	0	0.976	-0.976	–	–	Yes
11	XI	GPC	0.278	1.535	0	-0.754	0.754	–	–	Yes
12	XI	GPC	0.478	0.139	0	-0.219	0.219	–	–	No
13	XI	GPC	0.322	1.326	0	0.480	-0.480	–	–	Yes
14	XI	GPC	0.492	0.362	0	-0.388	0.388	–	–	No
15	XI	GPC	0.619	0.256	0	1.163	-1.163	–	–	Yes
16	XI	GPC	0.292	-1.341	0	-4.294	4.294	–	–	Yes
17	XI	GPC	0.504	-0.427	0	-0.953	0.953	–	–	No
18	XI	GPC	0.400	0.801	0	0.402	-0.402	–	–	No
19	XI	GPC	0.595	-0.203	0	1.180	-1.180	–	–	No
20	XI	GPC	0.228	0.743	0	-0.286	0.286	–	–	No
21	XI	GPC	0.411	-0.581	0	-2.416	2.416	–	–	Yes
22	XI	GPC	0.519	-0.313	0	-1.019	1.019	–	–	Yes
23	XI	GPC	0.500	-0.026	0	0.621	-0.621	–	–	Yes
24	XI	GPC	0.345	0.211	0	-1.515	1.515	–	–	Yes
25	XI	GPC	0.488	-0.539	0	-1.568	1.568	–	–	Yes
26	XI	GPC	0.353	0.117	0	0.964	-0.964	–	–	Yes
27	XI	GPC	0.381	1.392	0	-0.477	0.477	–	–	No

Table K.13. Operational Item Parameter Estimates—CSLA Grade 3

Item	Item Type	Model	B	D1	D2	D3	D4	Infit	Outfit
1	CR	Rasch	0.267	0	-0.866	0.150	0.716	0.89	0.86
2	CR	Rasch	0.297	0	-0.579	-0.017	0.596	0.82	0.75
3	CR	Rasch	0.182	0	-0.880	0.688	0.192	0.92	0.9
4	CR	Rasch	0.761	0	-0.729	-0.080	0.808	0.84	0.77
5	XI	Rasch	-1.158	0	-0.567	0.567	–	0.83	0.81
6	XI	Rasch	-0.004	0	0.040	-0.040	–	1.04	1.07
7	XI	Rasch	0.070	0	0.437	-0.437	–	1.27	1.42
8	XI	Rasch	-0.706	0	0.436	-0.436	–	0.86	0.84
9	XI	Rasch	-0.105	0	0.276	-0.276	–	0.93	0.89
10	XI	Rasch	0.450	0	-0.180	0.180	–	1.24	1.36
11	XI	Rasch	-0.329	0	0.788	-0.788	–	0.92	0.98
12	XI	Rasch	-0.862	0	-0.174	0.174	–	0.86	0.85
13	XI	Rasch	0.157	0	-0.022	0.022	–	1.08	1.13
14	XI	Rasch	-0.062	0	0.539	-0.539	–	1.23	1.37
15	XI	Rasch	-0.387	0	0.805	-0.805	–	1.02	1.13
16	XI	Rasch	0.052	0	0.900	-0.900	–	1.05	1.02
17	XI	Rasch	-0.306	0	-0.033	0.033	–	0.87	0.85
18	XI	Rasch	0.157	0	-0.090	0.090	–	1.10	1.25
19	XI	Rasch	0.351	0	0.552	-0.552	–	1.24	1.47
20	XI	Rasch	0.094	0	-0.017	0.017	–	1.07	1.13
21	XI	Rasch	0.030	0	0.338	-0.338	–	0.86	0.85
22	XI	Rasch	-0.034	0	0.614	-0.614	–	0.93	0.86
23	XI	Rasch	1.272	0	0.751	-0.751	–	1.17	1.16

Table K.14. Operational Item Parameter Estimates—CSLA Grade 4

Item	Item Type	Model	B	D1	D2	D3	D4	D5	Infit	Outfit
1	CR	Rasch	-0.227	0	-0.754	0.613	0.142	–	0.75	0.73
2	CR	Rasch	0.302	0	-0.722	-0.349	0.012	1.059	0.60	0.61
3	CR	Rasch	-0.055	0	-0.942	0.709	0.233	–	0.98	0.97
4	CR	Rasch	-0.374	0	-0.048	-0.389	0.437	–	0.93	0.86
5	XI	Rasch	-1.463	0	-0.205	0.205	–	–	0.87	0.87
6	XI	Rasch	0.792	0	-0.786	0.786	–	–	0.82	0.91
7	XI	Rasch	-0.991	0	0.699	-0.699	–	–	0.90	0.86
8	XI	Rasch	-0.149	0	0.836	-0.836	–	–	0.95	1.02
9	XI	Rasch	-1.225	0	0.582	-0.582	–	–	0.86	0.78
10	XI	Rasch	-1.057	0	0.664	-0.664	–	–	0.98	0.98
11	XI	Rasch	-1.185	0	-0.039	0.039	–	–	0.80	0.78
12	XI	Rasch	-1.120	0	0.699	-0.699	–	–	0.84	0.83
13	XI	Rasch	-1.410	0	0.600	-0.600	–	–	1.01	1.13
14	XI	Rasch	-0.250	0	-0.523	0.523	–	–	0.97	1.02
15	XI	Rasch	0.248	0	0.811	-0.811	–	–	1.17	1.78
16	XI	Rasch	-0.006	0	1.434	-1.434	–	–	1.07	1.40
17	XI	Rasch	-0.320	0	0.227	-0.227	–	–	0.94	0.97
18	XI	Rasch	0.352	0	0.327	-0.327	–	–	1.39	1.95
19	XI	Rasch	0.854	0	-0.174	0.174	–	–	1.22	1.75
20	XI	Rasch	-0.704	0	0.332	-0.332	–	–	1.11	1.14
21	XI	Rasch	0.789	0	-0.157	0.157	–	–	1.06	1.27
22	XI	Rasch	-0.875	0	-0.335	0.335	–	–	0.88	0.86
23	XI	Rasch	-0.197	0	-0.162	0.162	–	–	0.94	0.92
24	XI	Rasch	0.354	0	0.547	-0.547	–	–	0.90	0.87
25	XI	Rasch	0.402	0	-0.579	0.579	–	–	1.13	1.26

Table K.15. Operational Item Parameter Estimates—Science Grade 5

Item	Item Type	Model	A	B	C	D1	D2	D3	Misfit Flag
1	CR	GPC	0.667	0.726	–	0	0.176	-0.176	No
2	CR	GPC	0.649	1.120	–	0	0.989	-0.989	No
3	CR	GPC	0.678	0.818	–	0	0.951	-0.951	No
4	CR	GPC	0.933	0.280	–	0	0.461	-0.461	Yes
5	CR	GPC	1.020	0.628	–	0	0.113	-0.113	No
6	CR	GPC	0.668	0.161	–	0	-0.301	0.301	No
7	CR	GPC	0.669	-0.347	–	0	0.560	-0.560	No
8	CR	GPC	0.804	0.626	–	0	0.557	-0.557	No
9	CR	GPC	0.874	0.522	–	0	0.637	-0.637	No
10	CR	GPC	0.541	1.368	–	0	0.707	-0.707	No
11	CR	GPC	0.780	1.802	–	0	1.047	-1.047	No
12	CR	GPC	0.702	0.426	–	0	-0.479	0.479	No
13	XI	2PL	0.579	2.293	–	–	–	–	No
14	XI	2PL	0.709	0.259	–	–	–	–	No
15	XI	2PL	0.332	-0.079	–	–	–	–	No
16	XI	2PL	0.600	-0.200	–	–	–	–	No
17	XI	2PL	0.612	-0.974	–	–	–	–	No
18	XI	3PL	1.113	0.294	0.022	–	–	–	No
19	XI	3PL	1.146	1.541	0.129	–	–	–	No
20	XI	3PL	0.363	0.212	0.068	–	–	–	No
21	XI	3PL	0.682	0.203	0.104	–	–	–	No
22	XI	3PL	1.929	1.623	0.076	–	–	–	No
23	XI	3PL	0.614	-0.160	0.031	–	–	–	No
24	XI	3PL	0.821	1.586	0.122	–	–	–	No
25	XI	3PL	0.878	-0.739	0.085	–	–	–	No
26	XI	3PL	1.050	2.029	0.231	–	–	–	No
27	XI	3PL	0.621	0.967	0.043	–	–	–	No
28	XI	3PL	0.745	0.641	0.151	–	–	–	No
29	XI	3PL	0.803	-1.692	0.019	–	–	–	No
30	SR	3PL	0.787	2.317	0.294	–	–	–	No
31	SR	3PL	0.776	1.426	0.364	–	–	–	No
32	SR	3PL	0.613	2.109	0.159	–	–	–	No
33	SR	3PL	1.306	0.738	0.394	–	–	–	No
34	SR	3PL	0.838	-0.213	0.031	–	–	–	No
35	SR	3PL	0.447	-0.023	0.111	–	–	–	No
36	SR	3PL	0.685	-0.537	0.179	–	–	–	No
37	SR	3PL	0.836	0.486	0.235	–	–	–	No
38	SR	3PL	1.004	-0.145	0.271	–	–	–	No
39	SR	3PL	0.585	0.533	0.214	–	–	–	No

Table K.16. Operational Item Parameter Estimates—Science Grade 8

Item	Item Type	Model	A	B	C	D1	D2	D3	Misfit Flag
1	CR	GPC	1.128	1.299	–	0	0.241	-0.241	No
2	CR	GPC	0.993	1.832	–	0	0.327	-0.327	No
3	CR	GPC	0.997	1.100	–	0	0.134	-0.134	No
4	CR	GPC	0.545	0.646	–	0	0.031	-0.031	No
5	CR	GPC	0.826	1.403	–	0	0.655	-0.655	No
6	CR	GPC	0.715	1.128	–	0	0.225	-0.225	No
7	CR	GPC	0.829	-0.030	–	0	-0.226	0.226	No
8	CR	GPC	0.840	0.391	–	0	-0.143	0.143	No
9	CR	GPC	0.938	0.420	–	0	0.327	-0.327	No
10	CR	GPC	0.685	1.374	–	0	-0.708	0.708	No
11	CR	GPC	0.860	1.827	–	0	0.253	-0.253	No
12	CR	GPC	0.846	1.206	–	0	0.506	-0.506	No
13	CR	GPC	0.744	1.069	–	0	0.516	-0.516	No
14	CR	GPC	1.027	0.785	–	0	0.048	-0.048	No
15	XI	2PL	0.489	-0.129	–	–	–	–	No
16	XI	2PL	0.735	-0.755	–	–	–	–	No
17	XI	2PL	0.597	0.212	–	–	–	–	No
18	XI	2PL	0.716	2.260	–	–	–	–	No
19	XI	3PL	0.662	1.210	0.058	–	–	–	No
20	XI	3PL	0.601	-0.275	0.085	–	–	–	No
21	XI	3PL	1.308	1.915	0.049	–	–	–	No
22	XI	3PL	0.984	0.838	0.047	–	–	–	No
23	XI	3PL	1.671	1.742	0.146	–	–	–	No
24	XI	3PL	0.573	-0.595	0.284	–	–	–	No
25	XI	3PL	0.875	0.709	0.115	–	–	–	No
26	XI	3PL	1.181	1.034	0.330	–	–	–	No
27	XI	3PL	1.345	1.239	0.232	–	–	–	No
28	XI	3PL	0.986	0.921	0.188	–	–	–	No
29	XI	3PL	0.782	-0.206	0.087	–	–	–	No
30	XI	3PL	0.651	1.360	0.136	–	–	–	No
31	XI	3PL	0.202	2.280	0.063	–	–	–	No
32	XI	3PL	0.612	1.176	0.030	–	–	–	No
33	XI	3PL	0.815	1.260	0.185	–	–	–	No
34	XI	3PL	1.015	1.337	0.109	–	–	–	No
35	XI	3PL	1.044	1.803	0.088	–	–	–	No
36	XI	3PL	0.908	1.402	0.199	–	–	–	No
37	SR	3PL	0.966	0.562	0.307	–	–	–	No
38	SR	3PL	0.712	0.427	0.211	–	–	–	No
39	SR	3PL	1.244	1.441	0.290	–	–	–	No
40	SR	3PL	0.828	1.561	0.080	–	–	–	No
41	SR	3PL	1.003	-0.580	0.217	–	–	–	No
42	SR	3PL	1.660	0.477	0.274	–	–	–	No
43	SR	3PL	1.128	0.471	0.123	–	–	–	No
44	SR	3PL	1.146	1.497	0.284	–	–	–	No
45	SR	3PL	0.835	0.640	0.237	–	–	–	No
46	SR	3PL	0.980	-0.922	0.025	–	–	–	No
47	SR	3PL	1.712	-0.031	0.414	–	–	–	No

Table K.17. Operational Item Parameter Estimates—Science Grade 11

Item	Item Type	Model	A	B	C	D1	D2	D3	Misfit Flag
1	CR	GPC	0.592	2.604	–	0	0.163	-0.163	No
2	CR	GPC	0.777	0.507	–	0	-0.056	0.056	No
3	CR	GPC	0.778	0.070	–	0	-0.063	0.063	No
4	CR	GPC	0.821	0.852	–	0	0.763	-0.763	No
5	CR	GPC	0.949	0.942	–	0	0.393	-0.393	No
6	CR	GPC	0.861	0.068	–	0	0.162	-0.162	No
7	CR	GPC	0.714	1.368	–	0	1.260	-1.260	No
8	CR	GPC	0.765	0.838	–	0	0.544	-0.544	No
9	CR	GPC	0.611	2.464	–	0	0.118	-0.118	No
10	CR	GPC	0.755	2.346	–	0	0.374	-0.374	No
11	XI	2PL	0.704	0.924	–	–	–	–	No
12	XI	3PL	0.799	0.289	0.087	–	–	–	No
13	XI	3PL	0.881	0.160	0.007	–	–	–	No
14	XI	3PL	0.335	0.628	0.016	–	–	–	No
15	XI	3PL	1.196	-0.169	0.097	–	–	–	No
16	XI	3PL	1.397	2.283	0.037	–	–	–	No
17	XI	3PL	1.499	-0.236	0.127	–	–	–	No
18	XI	3PL	0.720	0.618	0.086	–	–	–	No
19	XI	3PL	1.053	1.658	0.050	–	–	–	No
20	XI	3PL	0.757	0.271	0.060	–	–	–	No
21	SR	3PL	0.658	0.622	0.022	–	–	–	No
22	SR	3PL	1.697	1.850	0.285	–	–	–	No
23	SR	3PL	0.962	0.337	0.213	–	–	–	No
24	SR	3PL	0.594	1.207	0.081	–	–	–	No
25	SR	3PL	0.917	-0.680	0.370	–	–	–	No
26	SR	3PL	1.091	1.330	0.310	–	–	–	No
27	SR	3PL	0.687	0.549	0.195	–	–	–	No
28	SR	3PL	0.743	2.022	0.194	–	–	–	No
29	SR	3PL	0.484	0.020	0.131	–	–	–	No
30	SR	3PL	1.034	0.777	0.225	–	–	–	No
31	SR	3PL	1.041	1.051	0.233	–	–	–	No
32	SR	3PL	0.519	-0.191	0.128	–	–	–	No

Appendix L: TCC, TIC, and CSEM Curves

Figure L.1. Mathematics Grade 3 TCC

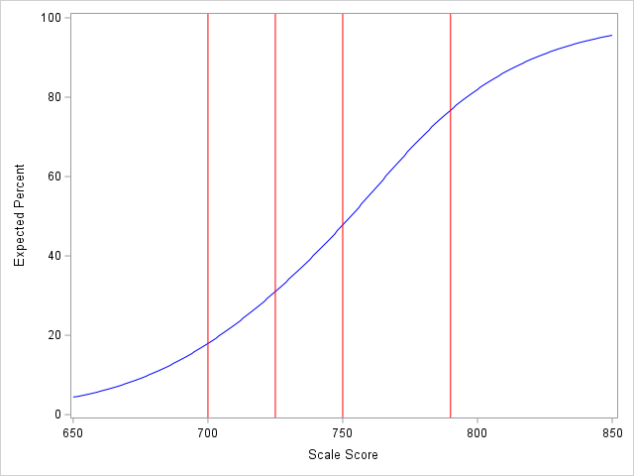


Figure L.2. Mathematics Grade 3 TIC

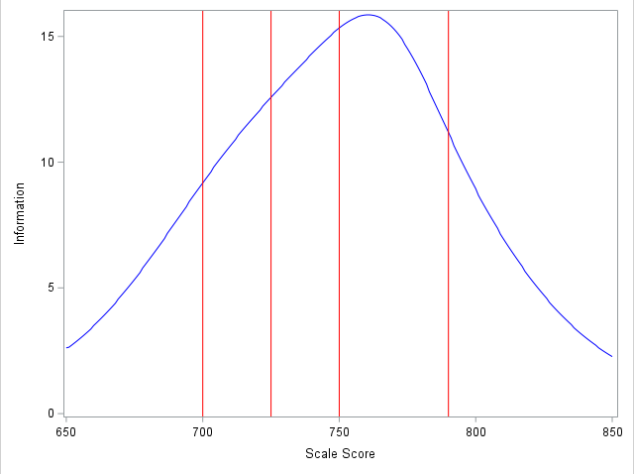


Figure L.3. Mathematics Grade 3 CSEM Curve

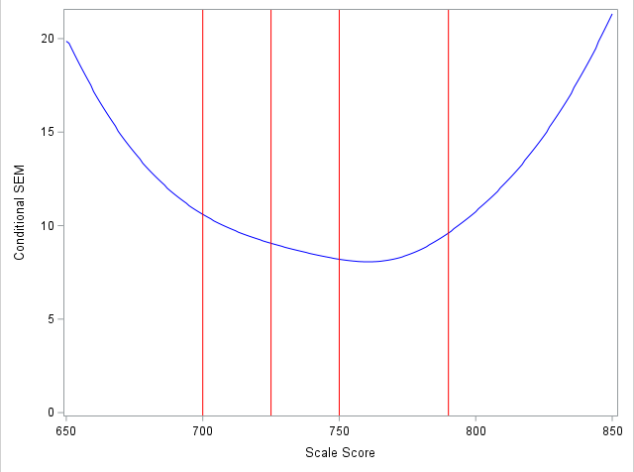


Figure L.4. Mathematics Grade 4 TCC

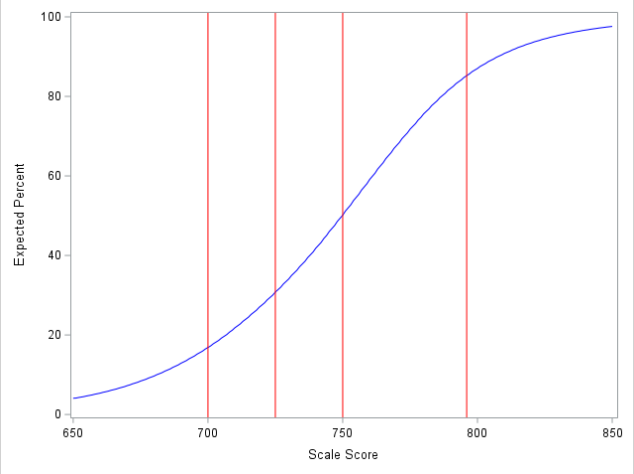


Figure L.5. Mathematics Grade 4 TIC

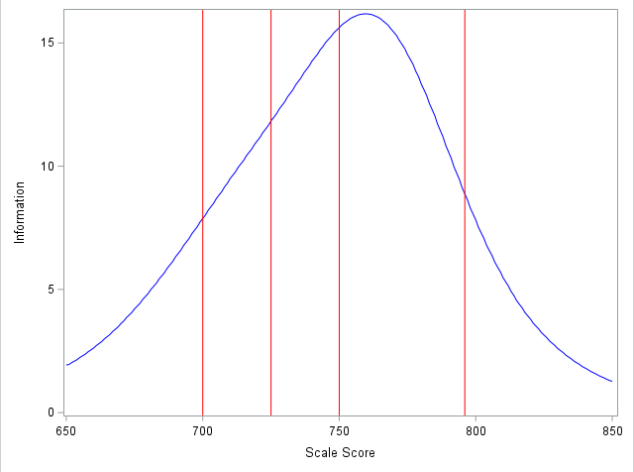


Figure L.6. Mathematics Grade 4 CSEM Curve

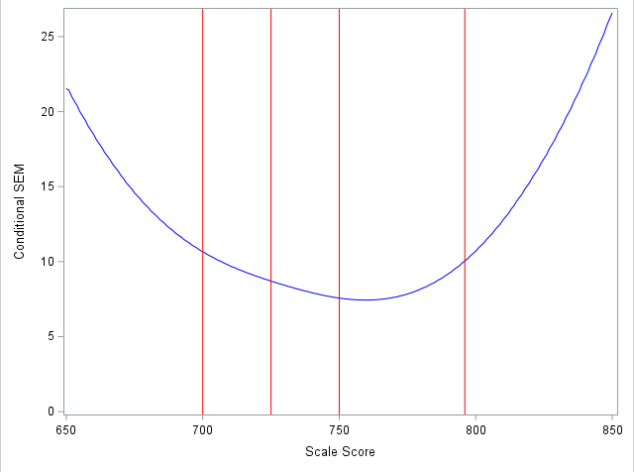


Figure L.7. Mathematics Grade 5 TCC

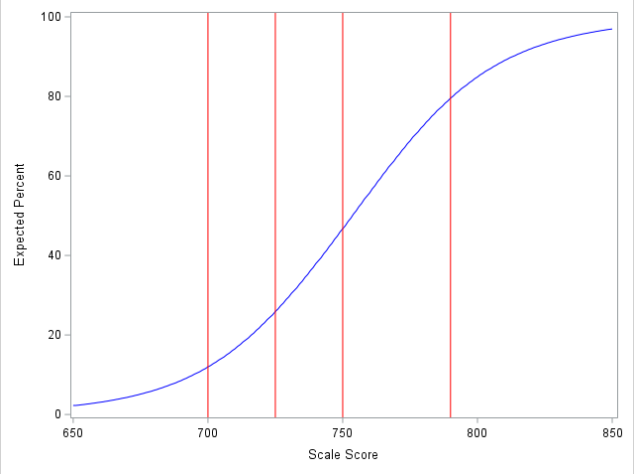


Figure L.8. Mathematics Grade 5 TIC

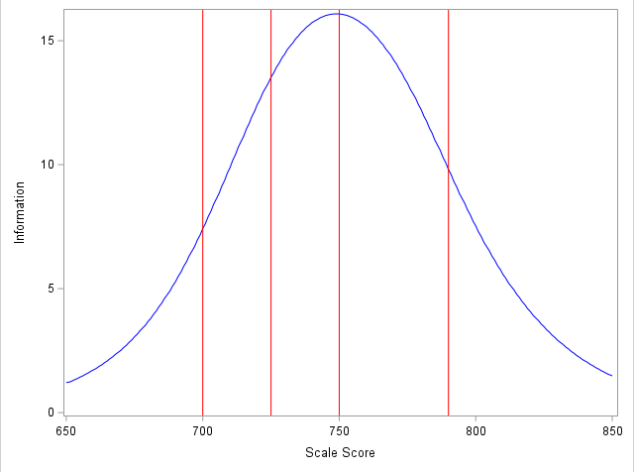


Figure L.9. Mathematics Grade 5 CSEM Curve

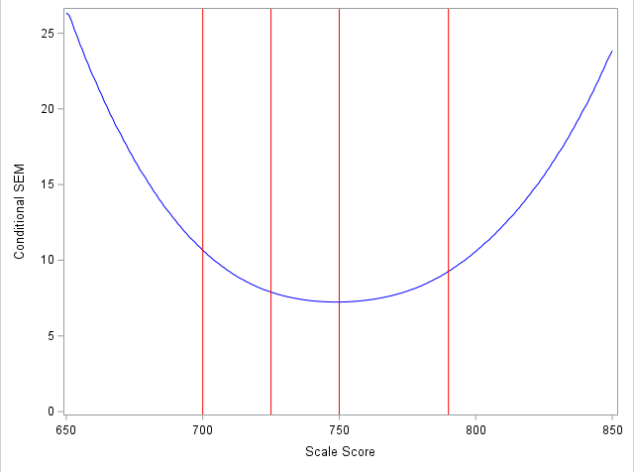


Figure L.10. Mathematics Grade 6 TCC

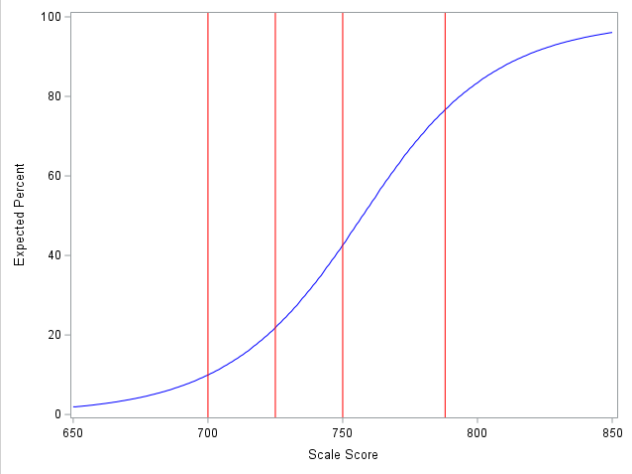


Figure L.11. Mathematics Grade 6 TIC

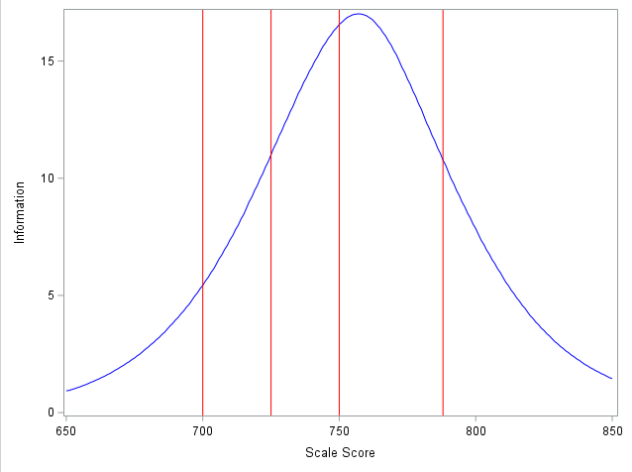


Figure L.12. Mathematics Grade 6 CSEM Curve

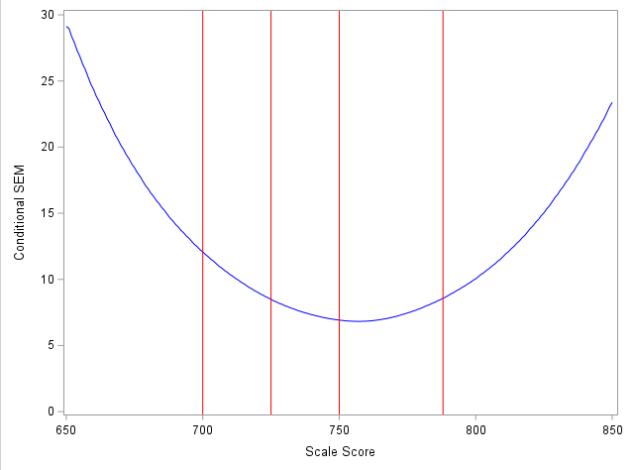


Figure L.13. Mathematics Grade 7 TCC

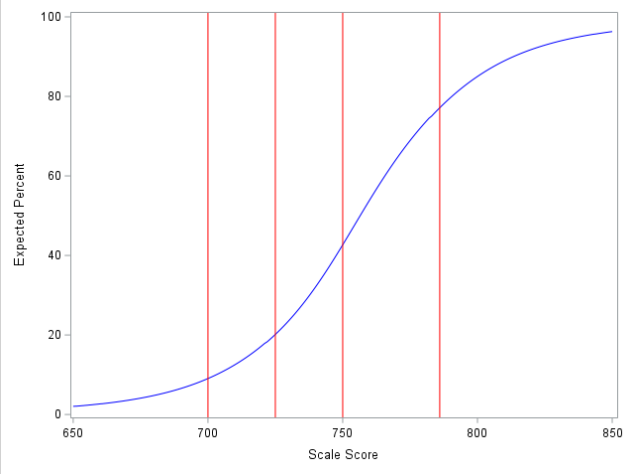


Figure L.14. Mathematics Grade 7 TIC

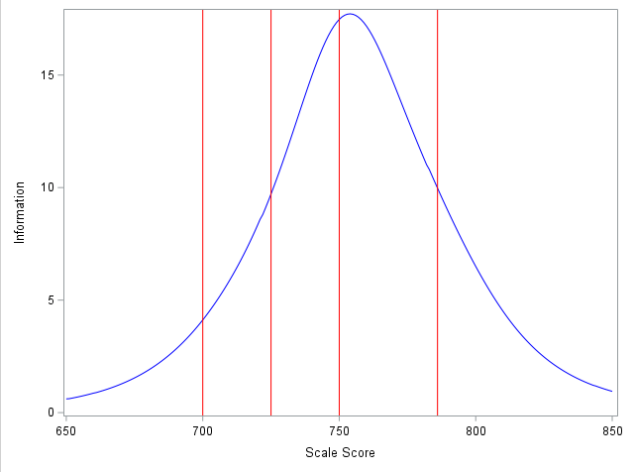


Figure L.15. Mathematics Grade 7 CSEM Curve

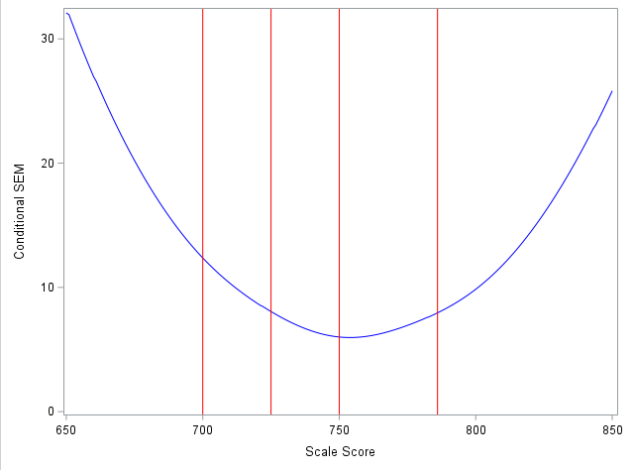


Figure L.16. Mathematics Grade 8 TCC

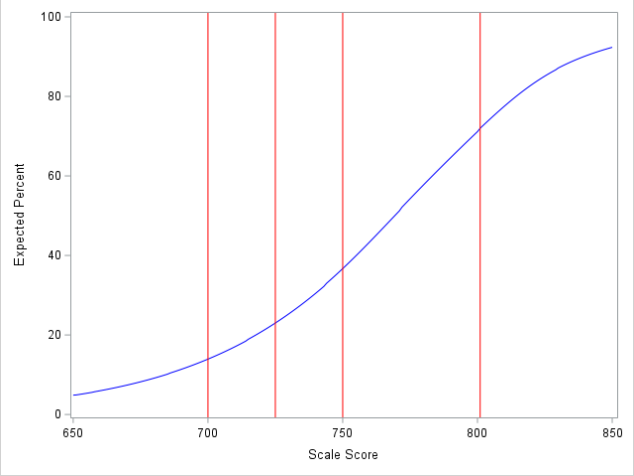


Figure L.17. Mathematics Grade 8 TIC

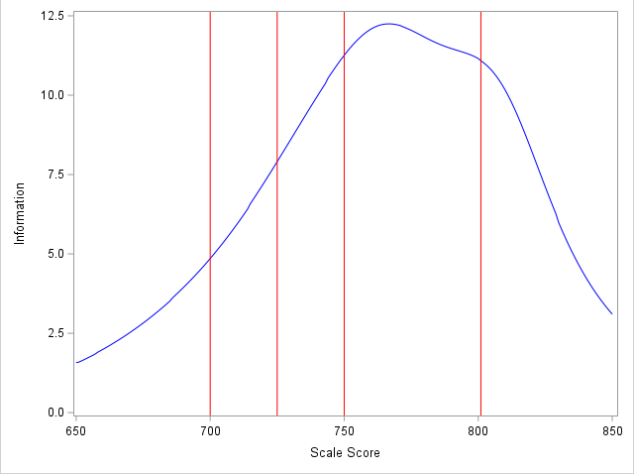


Figure L.18. Mathematics Grade 8 CSEM Curve

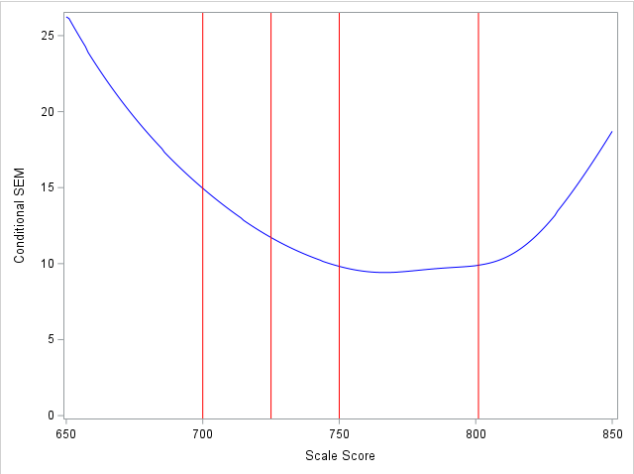


Figure L.19. ELA Grade 3 TCC

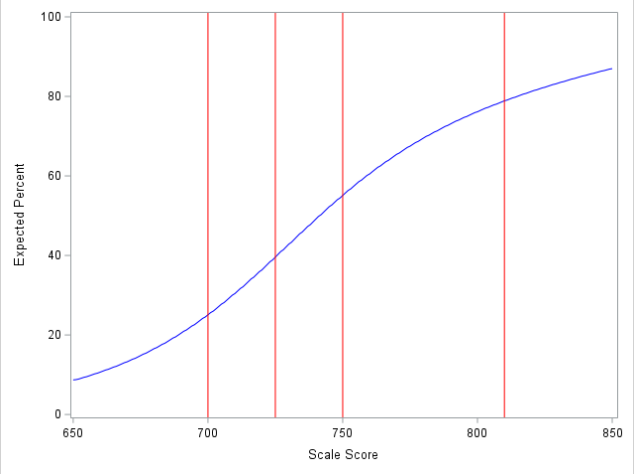


Figure L.20. ELA Grade 3 TIC

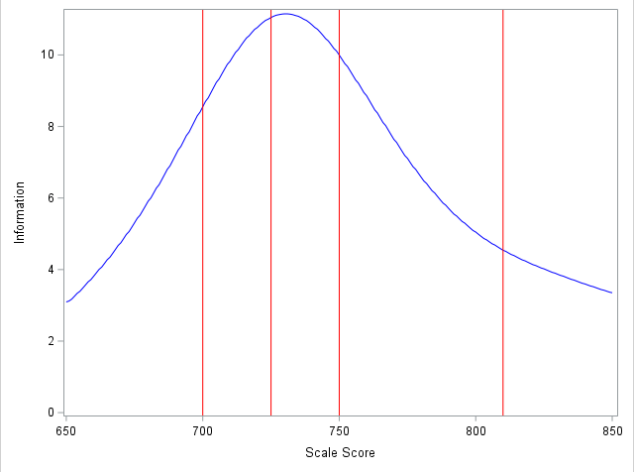


Figure L.21. ELA Grade 3 CSEM Curve

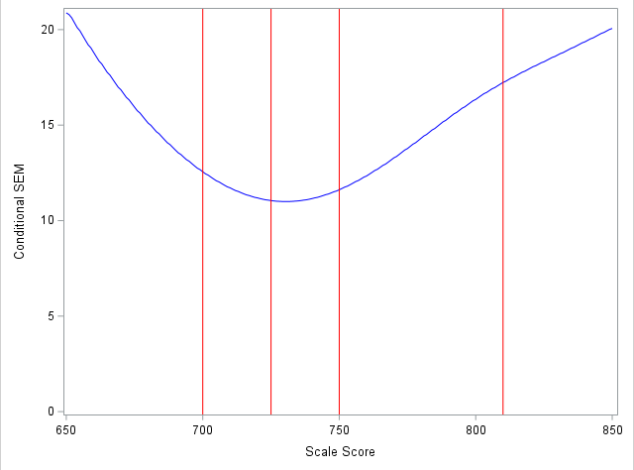


Figure L.22. ELA Grade 4 TCC

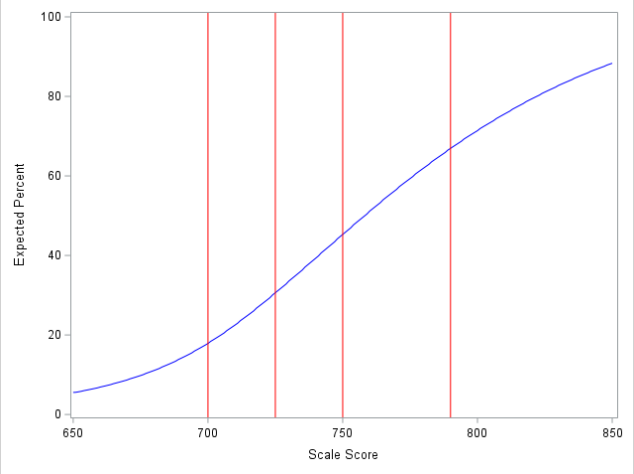


Figure L.23. ELA Grade 4 TIC

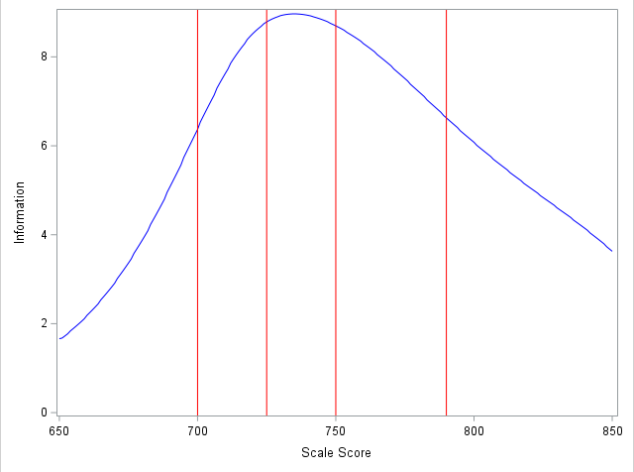


Figure L.24. ELA Grade 4 CSEM Curve

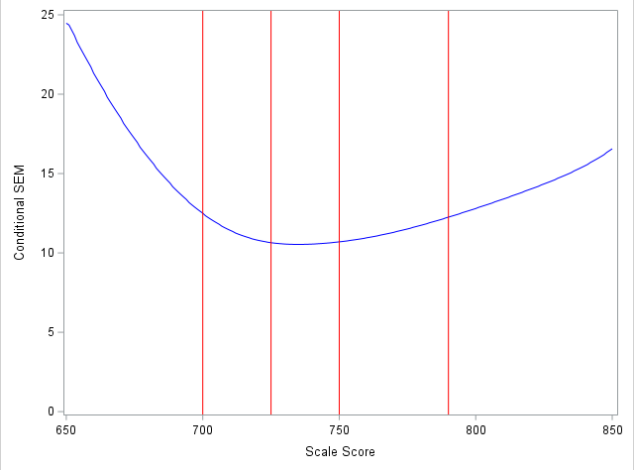


Figure L.25. ELA Grade 5 TCC

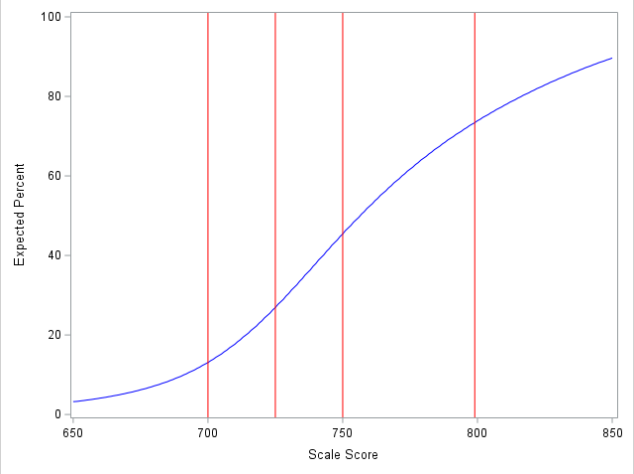


Figure L.26. ELA Grade 5 TIC

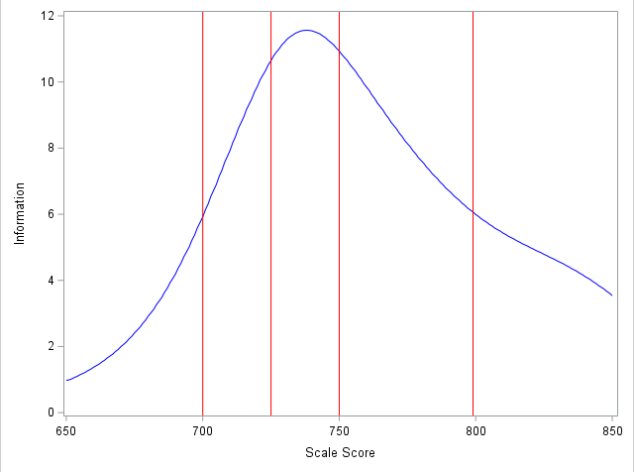


Figure L.27. ELA Grade 5 CSEM Curve

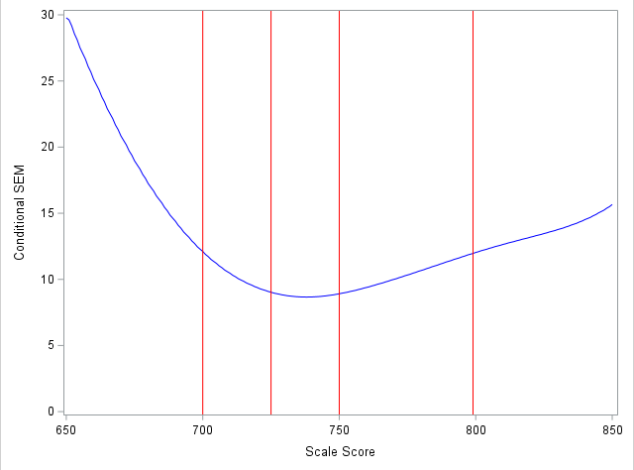


Figure L.28. ELA Grade 6 TCC

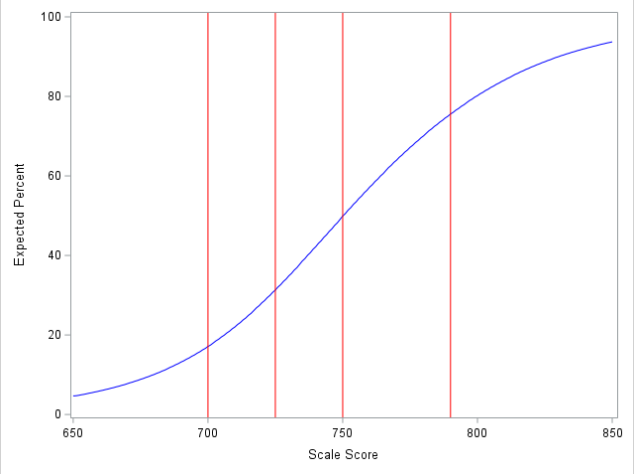


Figure L.29. ELA Grade 6 TIC

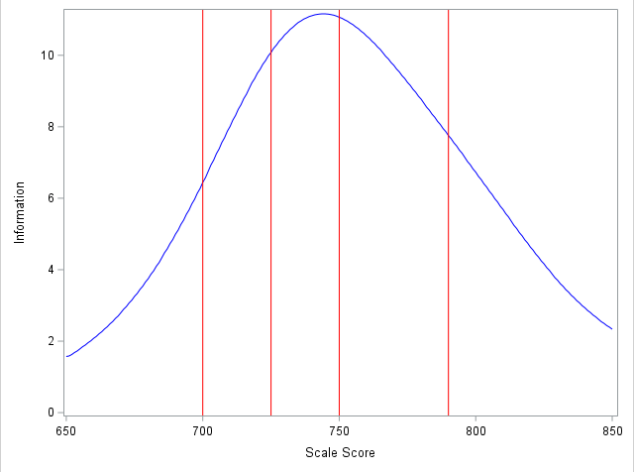


Figure L.30. ELA Grade 6 CSEM Curve

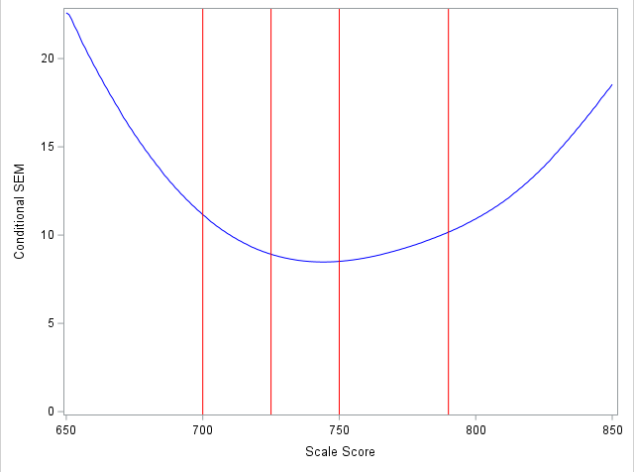


Figure L.31. ELA Grade 7 TCC

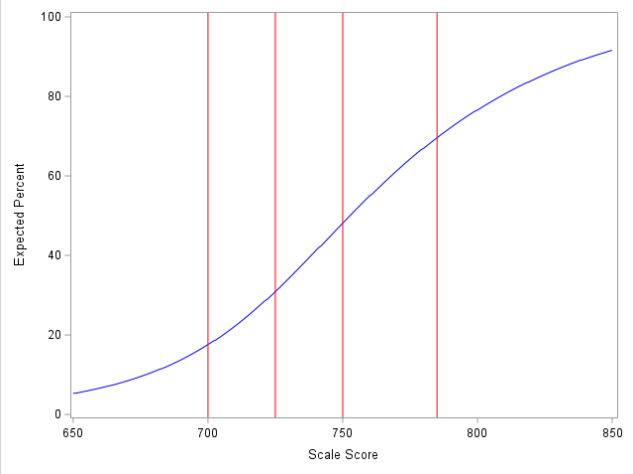


Figure L.32. ELA Grade 7 TIC

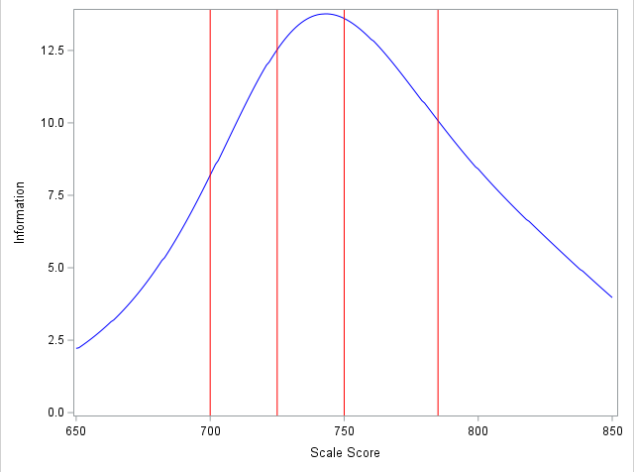


Figure L.33. ELA Grade 7 CSEM Curve

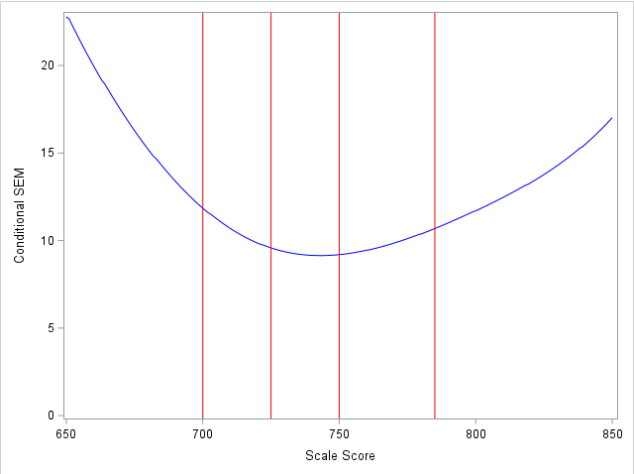


Figure L.34. ELA Grade 8 TCC

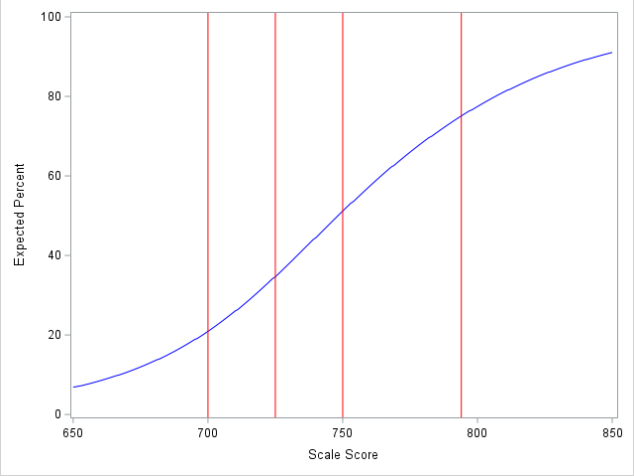


Figure L.35. ELA Grade 8 TIC

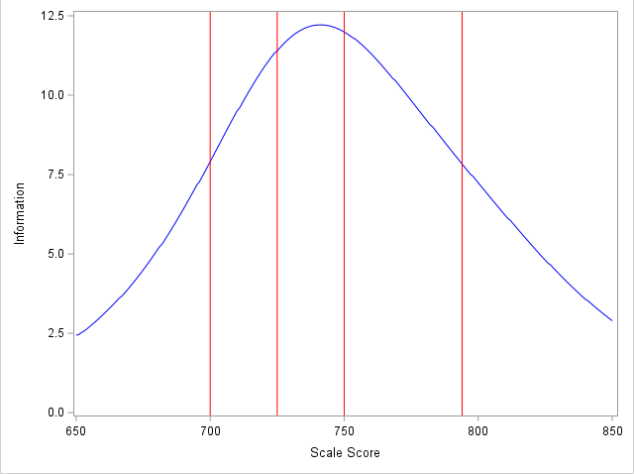


Figure L.36. ELA Grade 8 CSEM Curve

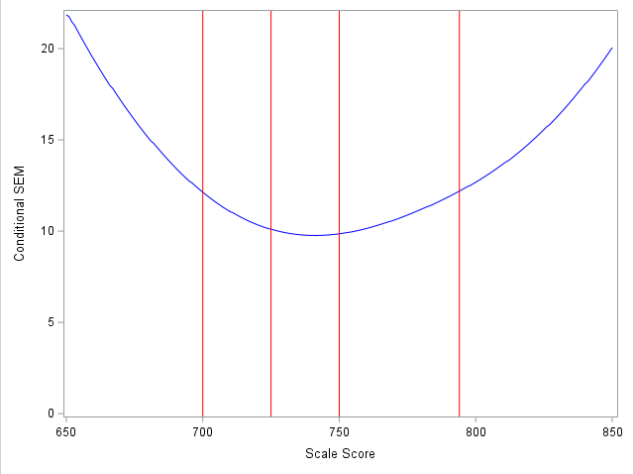


Figure L.37. CSLA Grade 3 TCC

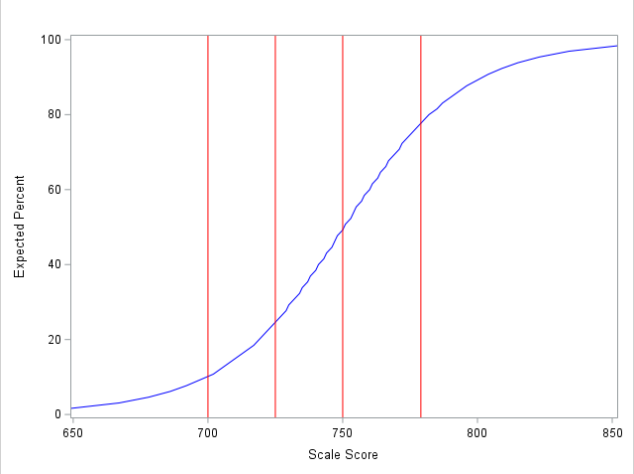


Figure L.38. CSLA Grade 3 TIC

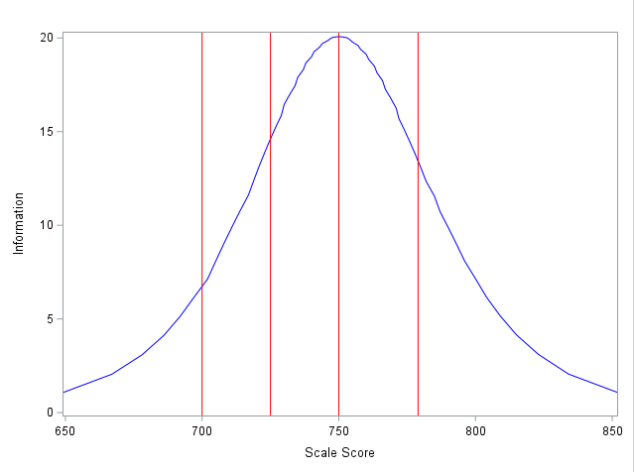


Figure L.39. CSLA Grade 3 CSEM Curve

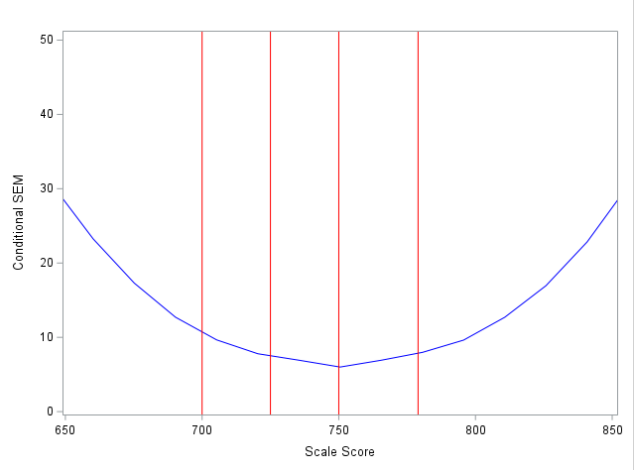


Figure L.40. CSLA Grade 4 TCC

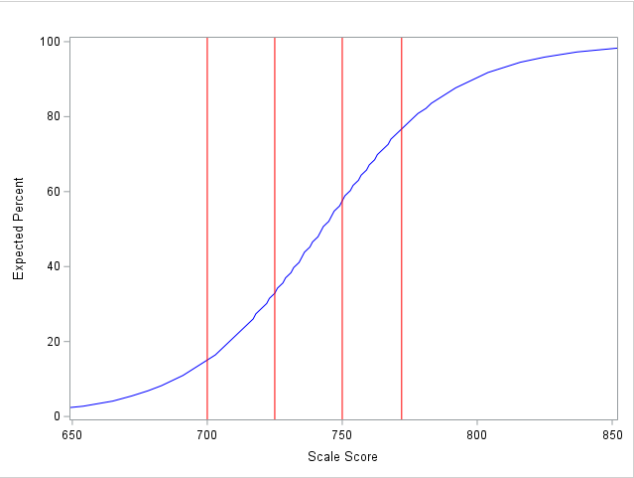


Figure L.41. CSLA Grade 4 TIC

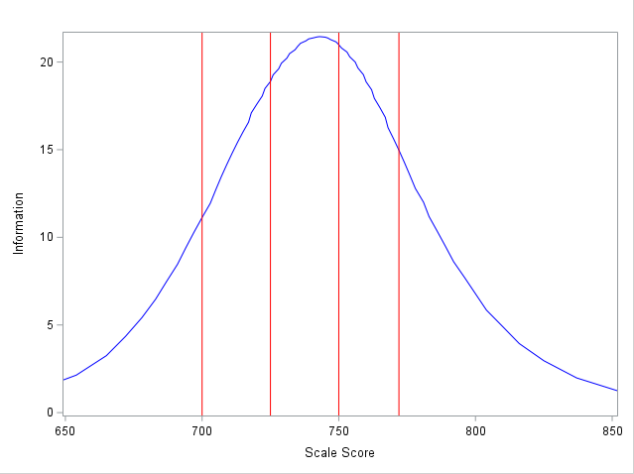


Figure L.42. CSLA Grade 4 CSEM Curve

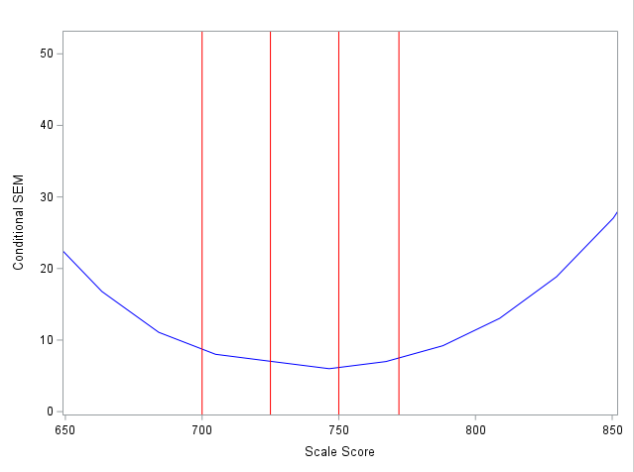


Figure L.43. ELA Reading Grade 3 TCC

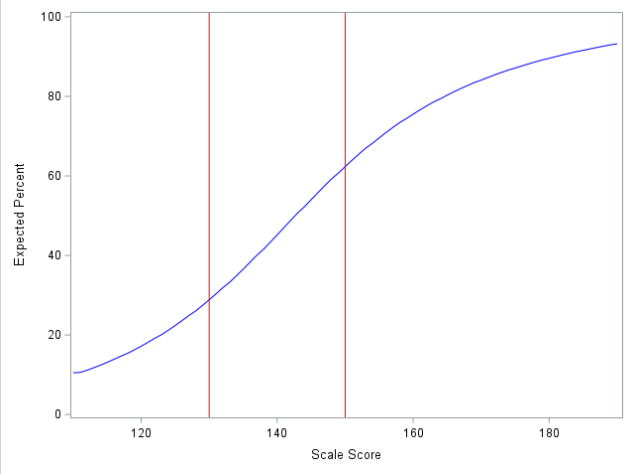


Figure L.44. ELA Reading Grade 3 TIC

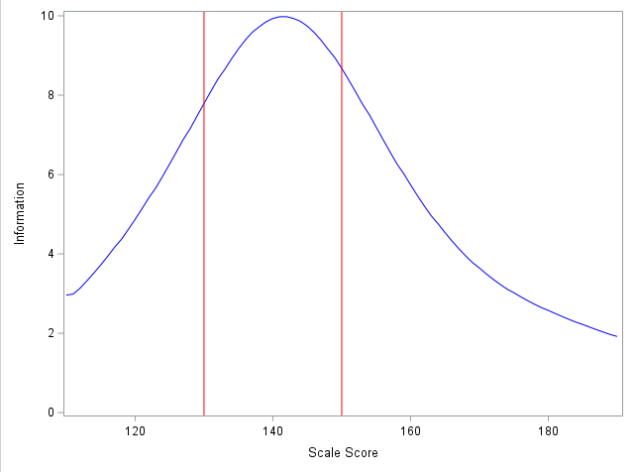


Figure L.45. ELA Reading Grade 3 CSEM Curve

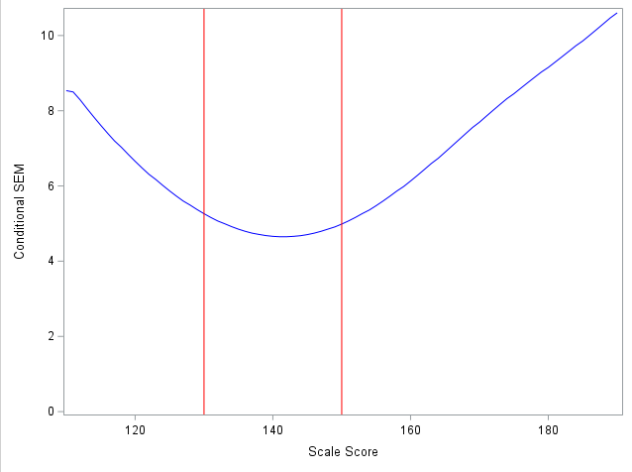


Figure L.46. ELA Reading Grade 4 TCC

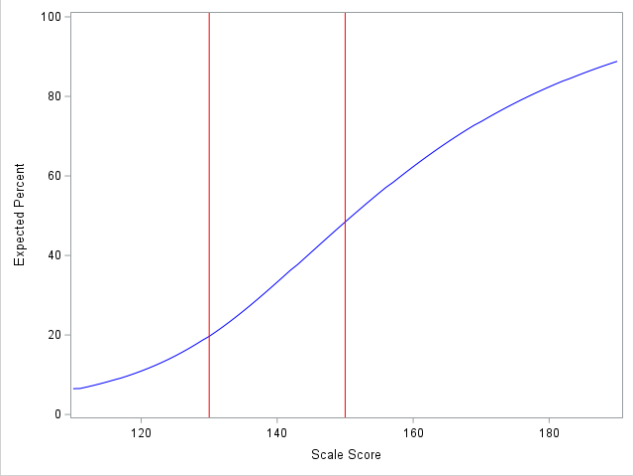


Figure L.47. ELA Reading Grade 4 TIC

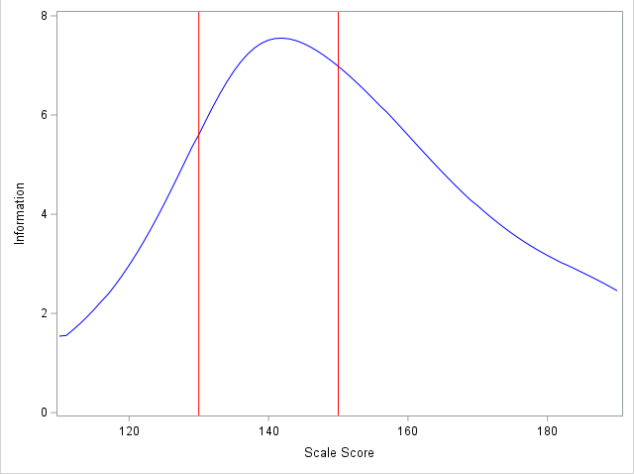


Figure L.48. ELA Reading Grade 4 CSEM Curve

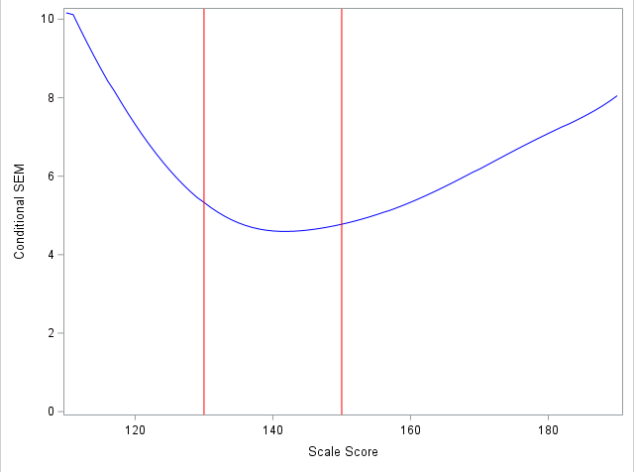


Figure L.49. ELA Reading Grade 5 TCC

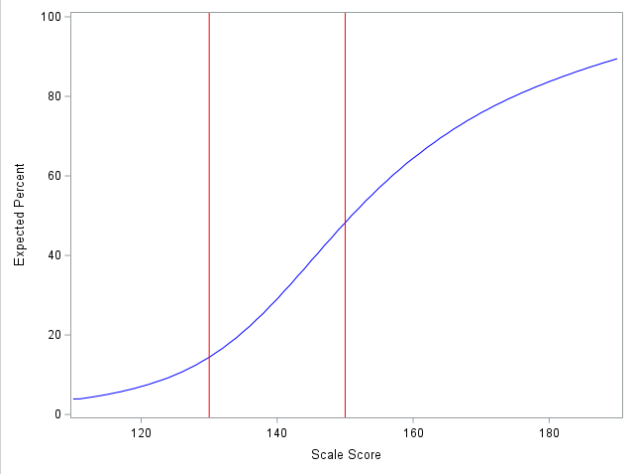


Figure L.50. ELA Reading Grade 5 TIC

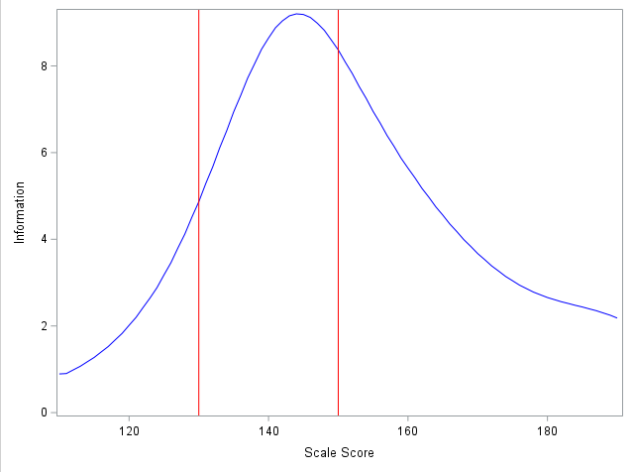


Figure L.51. ELA Reading Grade 5 CSEM Curve

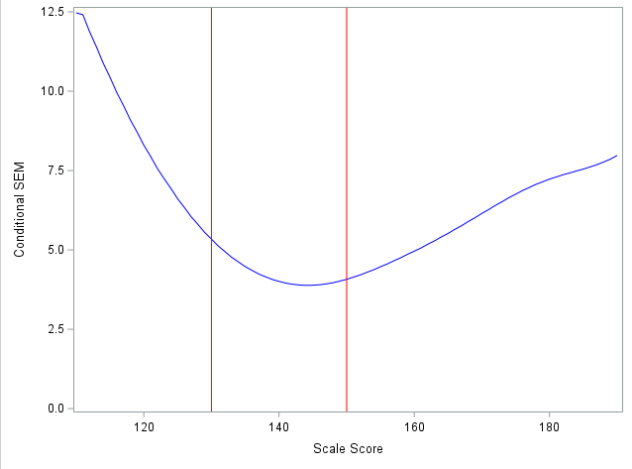


Figure L.52. ELA Reading Grade 6 TCC

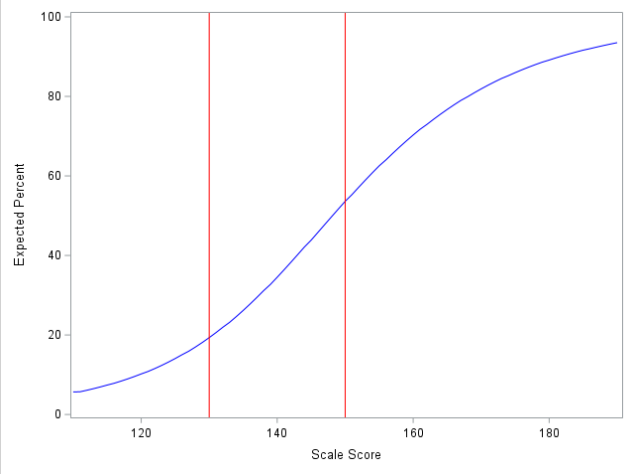


Figure L.53. ELA Reading Grade 6 TIC

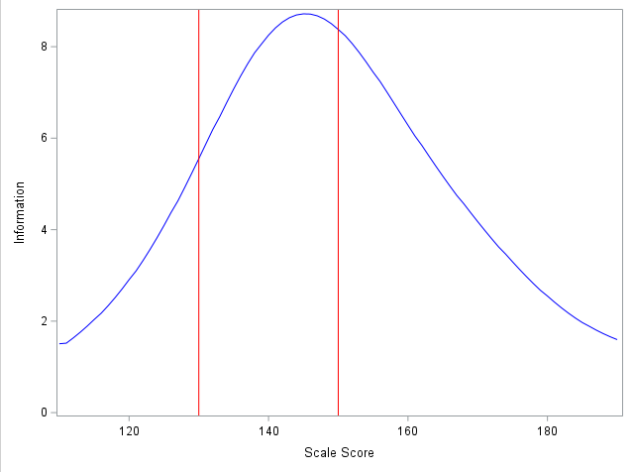


Figure L.54. ELA Reading Grade 6 CSEM Curve

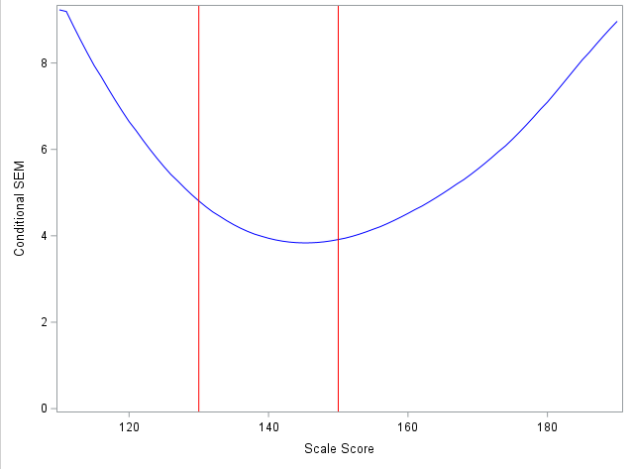


Figure L.55. ELA Reading Grade 7 TCC

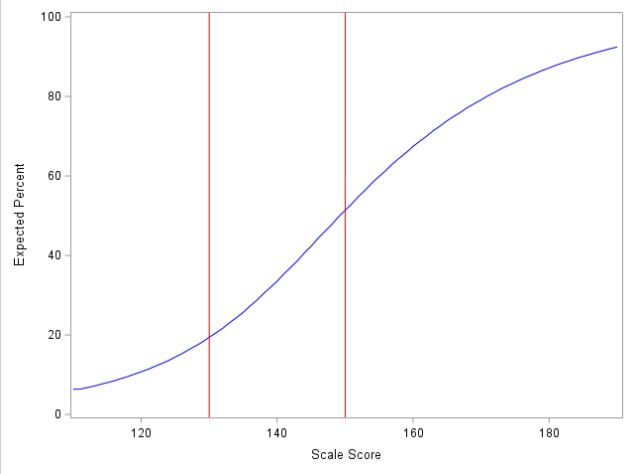


Figure L.56. ELA Reading Grade 7 TIC

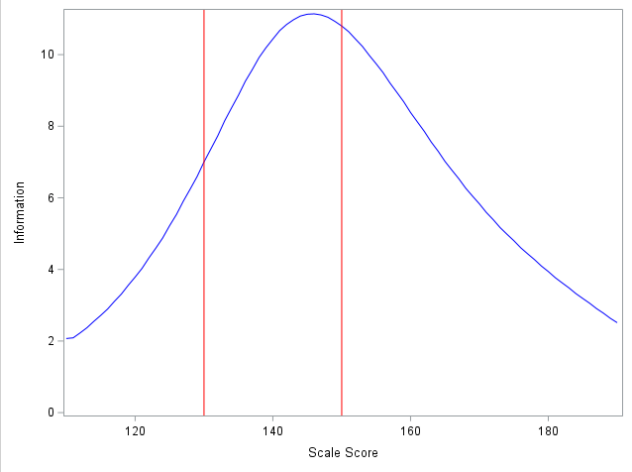


Figure L.57. ELA Reading Grade 7 CSEM Curve

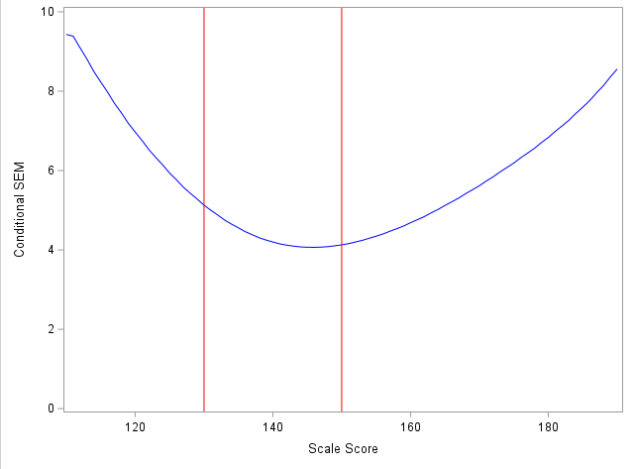


Figure L.58. ELA Reading Grade 8 TCC

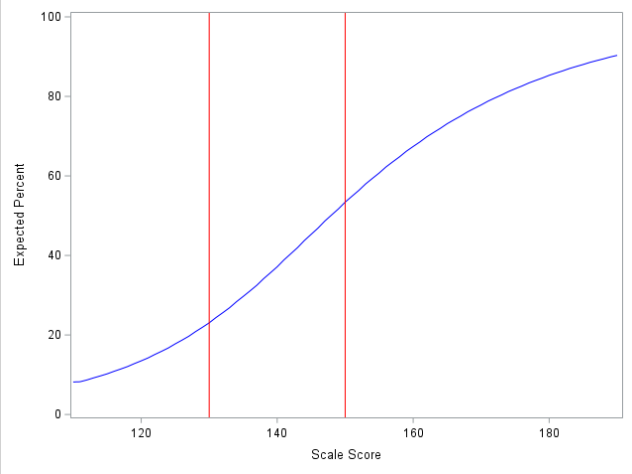


Figure L.59. ELA Reading Grade 8 TIC

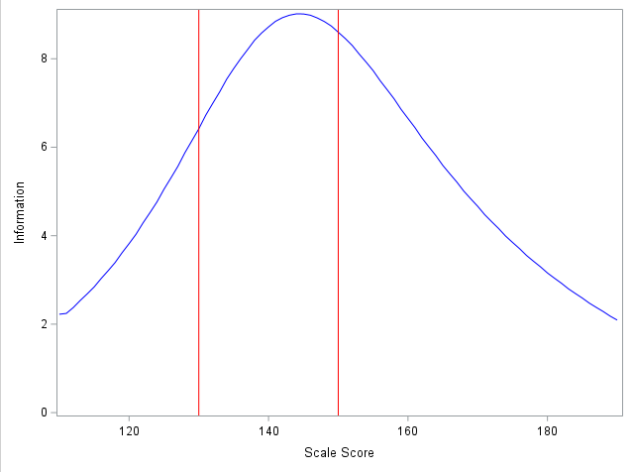


Figure L.60. ELA Reading Grade 8 CSEM Curve

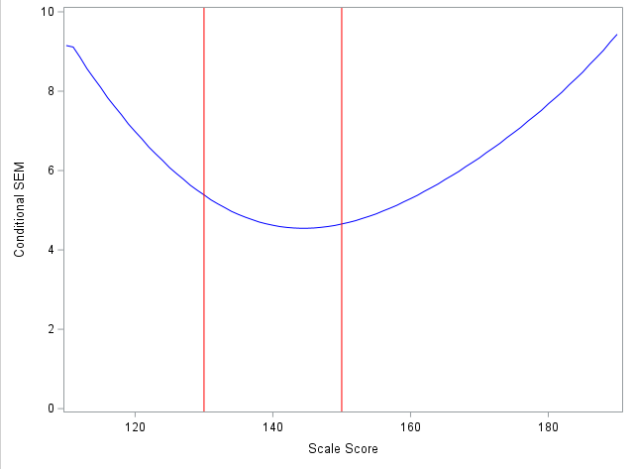


Figure L.61. CSLA Reading Grade 3 TCC

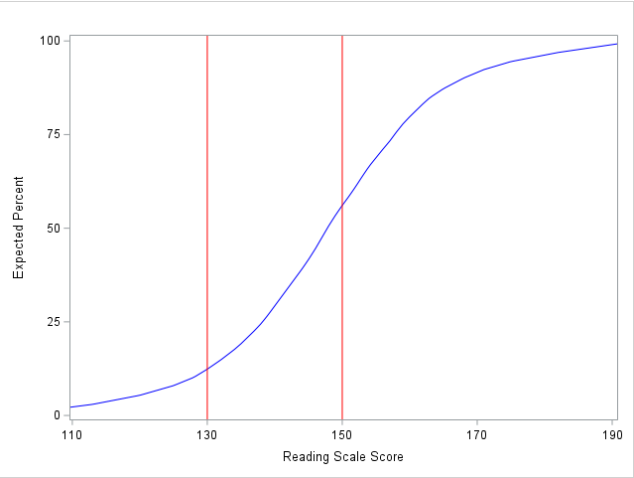


Figure L.62. CSLA Reading Grade 3 TIC

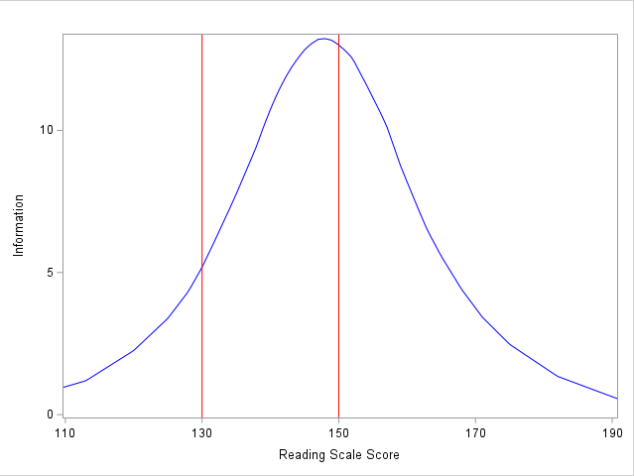


Figure L.63. CSLA Reading Grade 3 CSEM Curve

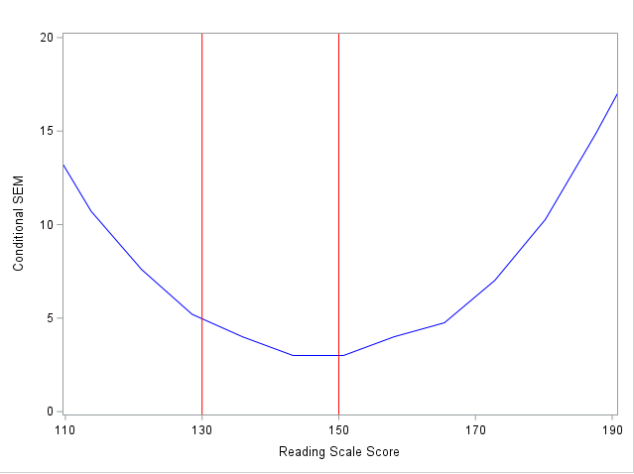


Figure L.64. CSLA Reading Grade 4 TCC

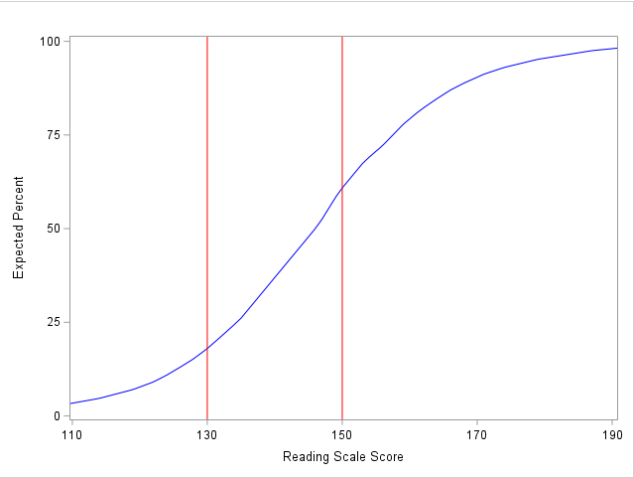


Figure L.65. CSLA Reading Grade 4 TIC

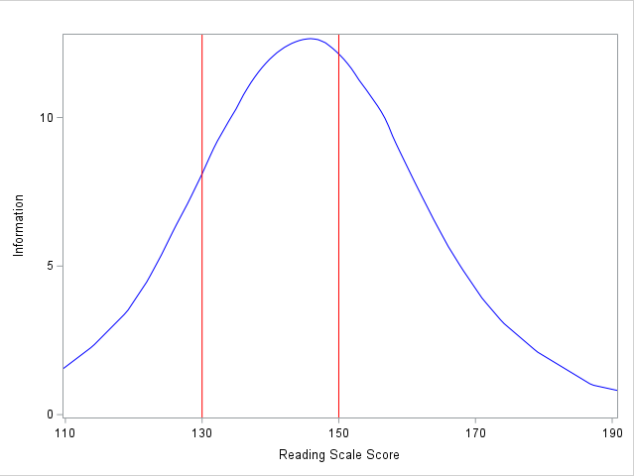


Figure L.66. CSLA Reading Grade 4 CSEM Curve

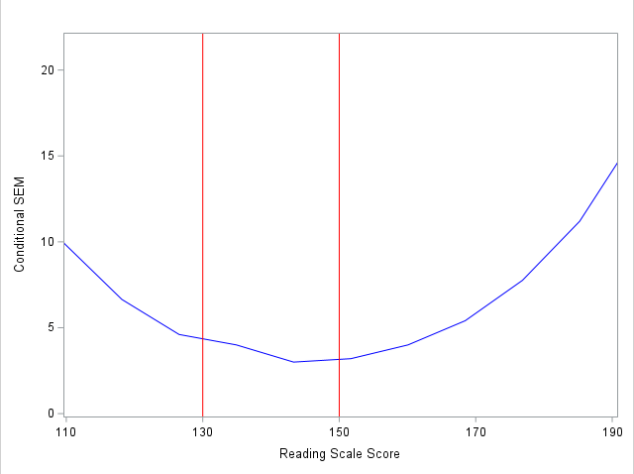


Figure L.67. Science Grade 5 TCC

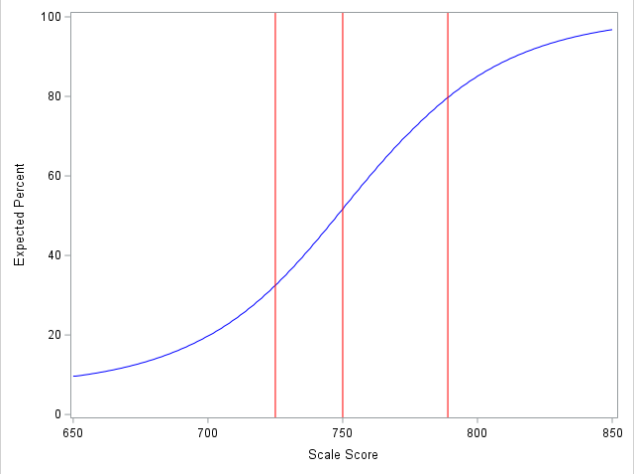


Figure L.68. Science Grade 5 TIC

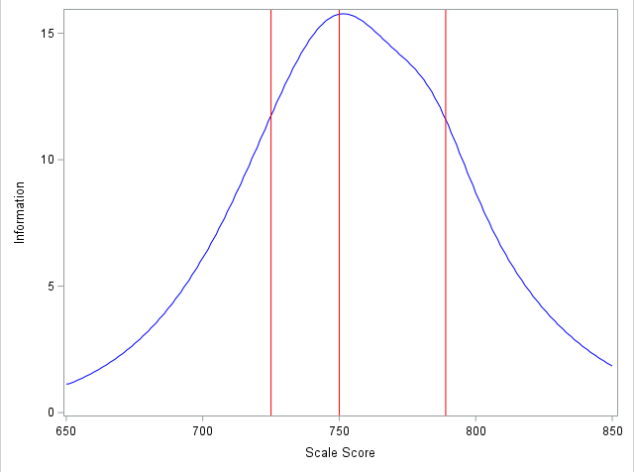


Figure L.69. Science Grade 5 CSEM

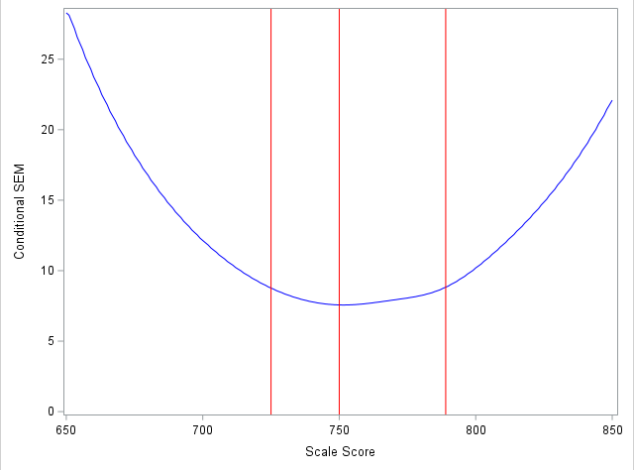


Figure L.70. Science Grade 8 TCC

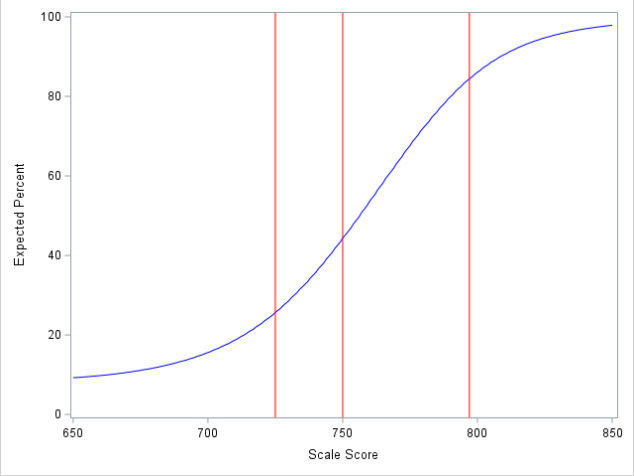


Figure L.71. Science Grade 8 TIC

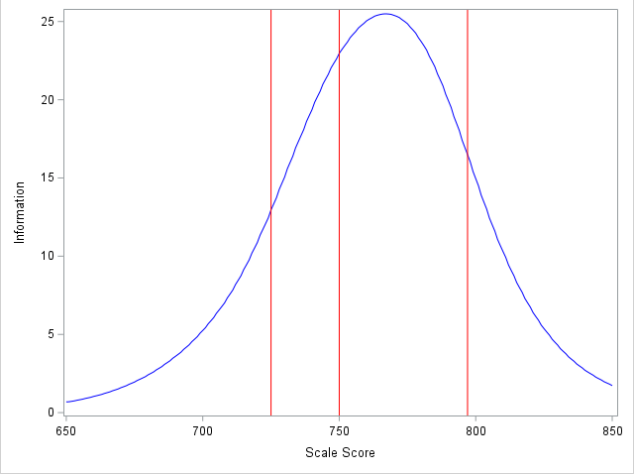


Figure L.72. Science Grade 8 CSEM

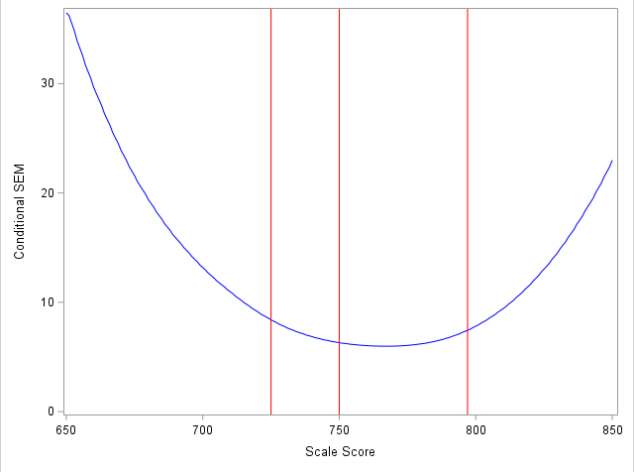


Figure L.73. Science Grade 11 TCC

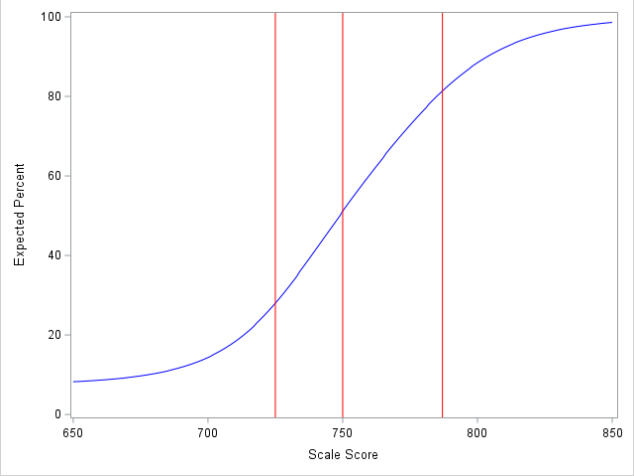


Figure L.74. Science Grade 11 TIC

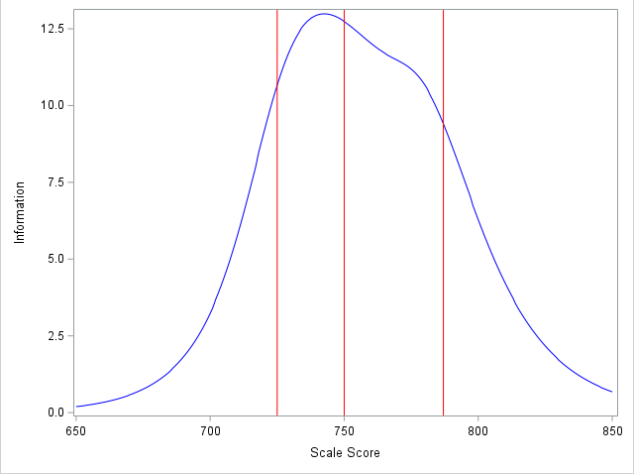
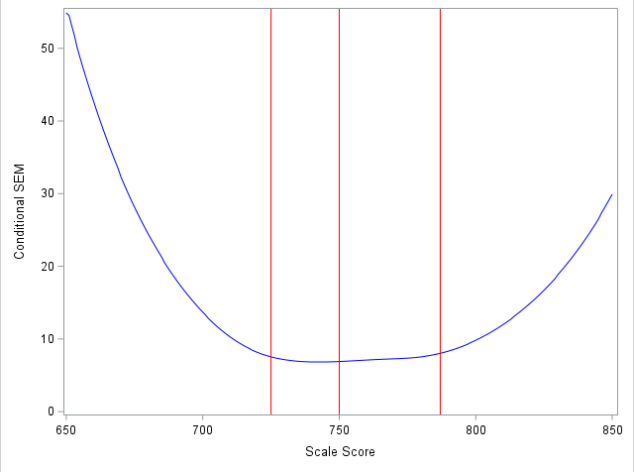


Figure L.75. Science Grade 11 CSEM



Appendix M: Inter-Rater Agreement

Table M.1. Operational Rater Agreement Statistics—Mathematics Grade 3

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	5,204	92.9	6.7	0.4	0.96	0.00	0.96
Item 2	4	5,408	85.1	14.1	0.8	0.90	0.01	0.90
Item 3	3	5,363	88.9	10.1	1.0	0.94	0.00	0.94
Item 4, PartB	3	5,003	85.1	14.2	0.7	0.93	0.00	0.93
Item 4, PartC	4	4,965	92.6	6.4	0.9	0.96	0.00	0.96

Table M.2. Operational Rater Agreement Statistics—Mathematics Grade 4

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, PartB	3	5,134	94.0	5.2	0.7	0.98	0.00	0.98
Item 2	3	5,502	87.2	12.3	0.5	0.95	0.00	0.95
Item 3, PartB	2	4,882	95.9	4.0	0.1	0.94	0.00	0.94
Item 3, PartA	2	5,336	89.7	10.0	0.3	0.91	0.01	0.91
Item 3, PartB	2	5,234	94.6	5.3	0.1	0.96	0.00	0.96
Item 3, PartB	2	5,295	94.8	4.9	0.3	0.93	0.00	0.93
Item 3, PartC	3	5,204	94.8	4.9	0.3	0.96	0.00	0.96

Table M.3. Operational Rater Agreement Statistics—Mathematics Grade 5

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, PartA	3	5,442	92.3	7.3	0.4	0.96	0.00	0.96
Item 1, PartB	3	5,194	93.4	5.4	1.3	0.95	0.01	0.95
Item 2	3	5,209	92.1	7.6	0.4	0.94	0.01	0.94
Item 3, PartB	3	5,179	86.4	13.1	0.6	0.89	0.01	0.89
Item 4	3	5,363	91.4	8.2	0.4	0.96	0.00	0.96
Item 5	4	5,430	86.7	12.4	0.9	0.95	0.00	0.95

Table M.4. Operational Rater Agreement Statistics—Mathematics Grade 6

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, PartA	4	5,271	90.4	6.9	2.7	0.94	0.00	0.94
Item 1, PartB	2	5,124	96.2	3.7	0.1	0.94	0.00	0.94
Item 2	3	5,321	87.5	12.0	0.5	0.93	0.01	0.93
Item 3	4	5,354	75.1	22.2	2.6	0.92	0.01	0.92
Item 4, PartB	2	5,124	92.2	7.4	0.4	0.88	0.00	0.88
Item 5	4	5,320	87.8	11.2	1.0	0.94	0.01	0.94

Table M.5. Operational Rater Agreement Statistics—Mathematics Grade 7

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, PartA	3	5,197	88.4	8.2	3.5	0.86	0.01	0.86
Item 1, PartB	3	5,087	91.9	6.3	1.8	0.95	0.00	0.95
Item 2	4	5,177	82.0	16.4	1.5	0.94	0.00	0.94
Item 3, PartB	2	4,749	93.0	6.7	0.4	0.81	0.01	0.81
Item 4	4	5,187	88.1	11.1	0.8	0.97	0.01	0.97
Item 5	3	5,099	89.1	9.6	1.3	0.90	0.00	0.90

Table M.6. Operational Rater Agreement Statistics—Mathematics Grade 8

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, PartA	3	4,674	97.0	2.6	0.4	0.94	0.00	0.94
Item 1, PartB	3	4,560	94.4	5.0	0.6	0.94	0.01	0.95
Item 2	3	4,654	95.5	4.1	0.3	0.92	0.01	0.92
Item 3, PartB	3	4,603	87.9	11.5	0.5	0.89	0.01	0.89
Item 4	3	4,908	89.0	10.2	0.8	0.94	0.00	0.94
Item 5, PartB	3	4,779	92.0	7.4	0.6	0.97	0.00	0.97

Table M.7. Field Test Rater Agreement Statistics—Mathematics Grade 3

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	2,998	82.1	17.1	0.7	0.86	0.01	0.86
Item 2, PartB	2	2,998	90.1	9.6	0.3	0.90	0.02	0.90
Item 3	3	2,998	87.8	11.5	0.7	0.94	0.00	0.94
Item 4	4	2,994	78.2	19.7	2.0	0.91	0.00	0.91
Item 5	3	2,999	82.7	16.2	1.1	0.91	0.02	0.91
Item 6	3	2,999	81.7	17.6	0.7	0.86	0.01	0.86

Table M.8. Field Test Rater Agreement Statistics—Mathematics Grade 4

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, PartB	2	2,990	96.2	3.7	0.0	0.97	0.00	0.97
Item 2	3	2,993	88.8	10.6	0.6	0.92	0.00	0.92
Item 3, PartB	3	2,998	92.8	7.1	0.1	0.97	0.00	0.97
Item 4	3	2,996	86.9	13.0	0.1	0.94	0.00	0.94
Item 5, PartA	2	2,998	91.2	8.8	0.0	0.93	0.01	0.93
Item 5, PartB	2	2,994	94.1	5.8	0.1	0.95	0.00	0.95
Item 5, PartA	2	2,998	93.1	6.8	0.0	0.95	0.00	0.95
Item 6	4	2,997	84.1	15.5	0.4	0.94	0.00	0.94

Table M.9. Field Test Rater Agreement Statistics—Mathematics Grade 5

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	2,995	83.6	11.6	4.8	0.86	0.01	0.86
Item 2	3	2,994	85.0	14.2	0.8	0.90	0.01	0.90
Item 3	3	2,990	80.3	18.6	1.1	0.89	0.01	0.89
Item 4	3	2,993	87.3	12.1	0.7	0.94	0.01	0.94
Item 5	4	2,993	81.7	17.1	1.1	0.91	0.00	0.91
Item 6	3	2,994	89.7	9.9	0.4	0.94	0.00	0.94
Item 7	3	2,978	90.0	9.2	0.8	0.94	0.00	0.94
Item 8, PartB	3	2,987	90.6	8.3	1.2	0.92	0.00	0.92
Item 8, PartC	2	2,986	90.6	9.0	0.4	0.93	0.00	0.93
Item 8, PartA	2	2,993	95.5	4.4	0.1	0.96	0.01	0.96
Item 9	4	2,996	76.3	20.3	3.3	0.90	0.02	0.90
Item 10	3	2,994	93.9	5.6	0.5	0.91	0.00	0.91
Item 11, PartC	2	2,993	79.6	19.4	1.0	0.80	0.01	0.80
Item 12	4	2,994	83.4	15.7	0.9	0.93	0.00	0.93
Item 13	3	2,995	87.8	11.4	0.8	0.95	0.00	0.95

Table M.10. Field Test Rater Agreement Statistics—Mathematics Grade 6

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	2,983	90.0	8.8	1.2	0.89	0.01	0.89
Item 2	3	2,990	93.0	7.0	0.1	0.96	0.00	0.96
Item 3	3	2,992	88.3	11.6	0.0	0.95	0.00	0.95
Item 4, PartB	3	2,975	92.5	7.5	0.0	0.92	0.00	0.92
Item 5	4	2,987	82.6	13.9	3.4	0.87	0.01	0.87
Item 6	4	2,993	93.7	5.5	0.8	0.98	0.00	0.98
Item 7	4	2,983	93.1	6.7	0.1	0.95	0.01	0.95
Item 8	3	2,983	96.5	3.5	0.0	0.98	0.00	0.98
Item 9, PartA	2	2,992	96.5	2.8	0.8	0.95	0.00	0.95
Item 9, PartB	2	2,988	95.0	4.7	0.2	0.88	0.00	0.88
Item 10	3	2,984	86.9	12.2	0.9	0.91	0.01	0.91

Table M.11. Field Test Rater Agreement Statistics—Mathematics Grade 7

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, PartB	2	2,984	95.7	4.2	0.1	0.94	0.00	0.94
Item 2	4	2,973	89.0	9.9	1.1	0.96	0.00	0.96
Item 3, PartA	2	2,989	89.6	8.9	1.5	0.91	0.00	0.91
Item 3, PartB	2	2,980	90.5	9.3	0.2	0.93	0.01	0.93
Item 3, PartB	2	2,980	95.6	4.2	0.2	0.96	0.00	0.96
Item 4	3	2,985	89.9	9.4	0.6	0.95	0.00	0.95
Item 5	3	2,979	91.7	7.9	0.4	0.95	0.00	0.95
Item 6	3	2,984	88.3	11.3	0.4	0.93	0.02	0.93
Item 7, PartB	3	2,974	92.6	6.8	0.6	0.95	0.00	0.95
Item 8	3	2,985	86.0	12.8	1.2	0.93	0.01	0.93
Item 9	3	2,980	81.4	16.4	2.1	0.86	0.01	0.86

Table M.12. Field Test Rater Agreement Statistics—Mathematics Grade 8

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, PartA	2	2,968	89.8	10.2	0.1	0.89	0.01	0.89
Item 2	3	2,975	92.3	7.2	0.5	0.93	0.00	0.93
Item 3	3	2,969	97.3	2.7	0.0	0.98	0.00	0.98
Item 4	4	2,968	80.8	18.2	0.9	0.92	0.01	0.92
Item 5	3	2,963	93.8	6.2	0.1	0.97	0.00	0.97
Item 6	3	2,975	95.3	4.4	0.3	0.97	0.01	0.97
Item 7	3	2,979	93.2	6.6	0.1	0.95	0.01	0.95
Item 8	3	2,965	93.9	5.9	0.2	0.95	0.01	0.95
Item 9	3	2,977	92.0	7.8	0.1	0.96	0.00	0.96
Item 10, PartA	3	2,986	88.6	10.7	0.7	0.96	0.00	0.96
Item 10, PartB	3	2,974	83.7	15.5	0.7	0.93	0.00	0.93

Table M.13. Operational Rater Agreement Statistics—ELA Grade 3

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	5,468	82.9	16.9	0.2	0.79	0.01	0.79
PCR 1 WE	3	5,468	85.5	14.5	0.0	0.79	0.01	0.79
PCR 2 WKL	3	5,462	85.1	14.7	0.2	0.77	0.00	0.77
PCR 2 WE	3	5,462	85.1	14.8	0.1	0.81	0.01	0.81

Table M.14. Operational Rater Agreement Statistics—ELA Grade 4

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	5,575	80.8	19.0	0.2	0.87	0.00	0.87
PCR 1 WE	3	5,575	80.4	19.1	0.5	0.85	0.03	0.85
PCR 2 WKL	4	5,573	76.7	23.2	0.1	0.75	0.00	0.75
PCR 2 WE	3	5,573	80.8	19.2	0.0	0.81	0.02	0.81

Table M.15. Operational Rater Agreement Statistics—ELA Grade 5

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	5,604	81.5	18.4	0.0	0.86	0.01	0.86
PCR 1 WE	3	5,604	81.8	18.2	0.1	0.86	0.00	0.86
PCR 2 WKL	4	5,598	76.6	23.1	0.3	0.78	0.02	0.78
PCR 2 WE	3	5,598	80.6	19.0	0.4	0.80	0.05	0.81

Table M.16. Operational Rater Agreement Statistics—ELA Grade 6

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	4	5,467	77.4	22.6	0.0	0.90	0.01	0.90
PCR 1 WE	3	5,467	80.9	19.1	0.0	0.90	0.01	0.90
PCR 2 WKL	4	5,453	78.9	20.8	0.4	0.81	0.01	0.81
PCR 2 WE	3	5,453	80.3	19.1	0.6	0.84	0.02	0.84

Table M.17. Operational Rater Agreement Statistics—ELA Grade 7

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	4	5,339	73.7	26.1	0.2	0.85	0.01	0.85
PCR 1 WE	3	5,339	79.4	20.2	0.3	0.88	0.04	0.88
PCR 2 WKL	4	5,351	72.9	27.0	0.1	0.84	0.00	0.84
PCR 2 WE	3	5,351	79.9	20.1	0.1	0.87	0.03	0.87

Table M.18. Operational Rater Agreement Statistics—ELA Grade 8

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	4	5,047	78.7	21.3	0.1	0.91	0.00	0.91
PCR 1 WE	3	5,047	82.8	17.1	0.0	0.92	0.00	0.92
PCR 2 WKL	4	5,040	75.4	24.4	0.2	0.90	0.01	0.90
PCR 2 WE	3	5,040	80.1	19.8	0.1	0.90	0.01	0.90

Table M.19. Field Test Rater Agreement Statistics—ELA Grade 3

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR_1_WE	3	3,000	73.2	26.8	0.0	0.73	0.03	0.73
PCR_1_WKL	3	3,000	83.9	16.1	0.0	0.82	0.00	0.82
PCR_2_WE	3	3,000	77.8	22.2	0.1	0.76	0.02	0.76
PCR_2_WKL	3	3,000	83.6	16.4	0.1	0.82	0.01	0.82
PCR_3_WE	3	3,000	95.1	4.7	0.2	0.88	0.01	0.88
PCR_3_WKL	3	3,000	94.3	5.6	0.1	0.92	0.00	0.92
PCR_4_WE	3	3,000	85.2	14.0	0.8	0.78	0.01	0.78
PCR_4_WKL	3	3,000	83.9	16.0	0.1	0.81	0.01	0.81

Table M.20. Field Test Rater Agreement Statistics—ELA Grade 4

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR_1_WE	3	2,999	80.9	19.1	0.0	0.89	0.01	0.89
PCR_1_WKL	3	2,999	81.7	18.3	0.0	0.88	0.02	0.88
PCR_2_WE	3	2,999	82.9	17.1	0.0	0.90	0.01	0.90
PCR_2_WKL	3	2,999	86.2	13.8	0.0	0.90	0.01	0.90
PCR_3_WE	4	2,996	77.0	23.0	0.0	0.79	0.00	0.79
PCR_3_WKL	3	2,996	80.1	19.8	0.1	0.80	0.02	0.81
PCR_4_WE	4	2,998	74.0	26.0	0.0	0.71	0.00	0.71
PCR_4_WKL	3	2,998	79.1	20.9	0.0	0.78	0.02	0.79
PCR_5_WE	4	2,999	75.8	24.2	0.0	0.79	0.02	0.79
PCR_5_WKL	3	2,999	80.6	19.3	0.1	0.82	0.00	0.82
PCR_6_WE	4	2,998	79.7	19.9	0.4	0.83	0.02	0.83
PCR_6_WKL	3	2,998	85.8	14.2	0.0	0.86	0.02	0.86

Table M.21. Field Test Rater Agreement Statistics—ELA Grade 5

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR_1_WE	4	2,999	88.1	11.7	0.2	0.92	0.02	0.92
PCR_1_WKL	3	2,999	87.1	12.4	0.5	0.87	0.01	0.87
PCR_2_WE	4	2,998	82.2	17.5	0.4	0.87	0.01	0.87
PCR_2_WKL	3	2,998	82.2	17.5	0.3	0.83	0.03	0.83
PCR_3_WE	4	2,999	76.2	23.5	0.3	0.80	0.01	0.80
PCR_3_WKL	3	2,999	79.9	19.8	0.3	0.83	0.02	0.83
PCR_4_WE	4	2,997	76.4	23.5	0.1	0.79	0.03	0.79
PCR_4_WKL	3	2,997	80.3	19.5	0.2	0.83	0.04	0.83
PCR_5_WE	3	2,998	90.1	9.9	0.1	0.94	0.01	0.94
PCR_5_WKL	3	2,998	88.4	11.5	0.1	0.91	0.00	0.91
PCR_6_WE	3	2,997	93.8	5.8	0.4	0.96	0.01	0.96
PCR_6_WKL	3	2,997	93.2	6.5	0.3	0.94	0.01	0.94
PCR_7_WE	3	2,998	80.2	19.6	0.2	0.89	0.03	0.89
PCR_7_WKL	3	2,998	80.0	19.8	0.2	0.87	0.04	0.87
PCR_8_WE	3	2,997	82.5	17.4	0.1	0.92	0.00	0.92
PCR_8_WKL	3	2,997	79.8	20.0	0.2	0.86	0.01	0.86

Table M.22. Field Test Rater Agreement Statistics—ELA Grade 6

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR_1_WE	4	2,979	80.8	18.9	0.3	0.86	0.00	0.86
PCR_1_WKL	3	2,979	79.3	20.1	0.6	0.86	0.01	0.86
PCR_2_WE	4	2,988	80.0	19.6	0.4	0.84	0.02	0.84
PCR_2_WKL	3	2,988	79.7	19.8	0.5	0.84	0.02	0.84
PCR_3_WE	4	2,989	81.8	17.9	0.3	0.85	0.01	0.85
PCR_3_WKL	3	2,989	81.3	18.1	0.6	0.87	0.03	0.87
PCR_4_WE	4	2,982	80.3	19.5	0.2	0.84	0.00	0.84
PCR_4_WKL	3	2,982	81.3	18.4	0.3	0.85	0.01	0.85

Table M.23. Field Test Rater Agreement Statistics—ELA Grade 7

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR_1_WE	4	2,985	78.5	21.2	0.3	0.90	0.02	0.90
PCR_1_WKL	3	2,985	82.2	17.6	0.2	0.91	0.01	0.91
PCR_2_WE	4	2,982	80.3	19.6	0.1	0.90	0.00	0.90
PCR_2_WKL	3	2,982	83.4	16.3	0.3	0.91	0.00	0.91
PCR_3_WE	4	2,985	77.0	23.0	0.0	0.88	0.01	0.88
PCR_3_WKL	3	2,985	78.4	21.6	0.0	0.88	0.02	0.88
PCR_4_WE	4	2,979	75.2	24.8	0.0	0.87	0.00	0.87
PCR_4_WKL	3	2,979	79.9	20.1	0.0	0.88	0.02	0.88

Table M.24. Field Test Rater Agreement Statistics—ELA Grade 8

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR_1_WE	4	2,991	87.6	12.4	0.0	0.94	0.00	0.94
PCR_1_WKL	3	2,991	85.1	14.9	0.0	0.93	0.00	0.93
PCR_2_WE	4	2,987	79.8	20.2	0.0	0.91	0.00	0.91
PCR_2_WKL	3	2,987	81.3	18.7	0.0	0.91	0.01	0.91

Table M.25. Operational Rater Agreement Statistics—CSLA Grade 3

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR_1_WE	3	158	90.6	9.4	0.0	0.95	0.04	0.95
PCR_1_WKL	3	158	87.9	12.1	0.0	0.93	0.01	0.93
PCR_2_WE	3	157	93.9	6.1	0.0	0.96	0.01	0.96
PCR_2_WKL	3	157	87.8	12.2	0.0	0.93	0.04	0.93

Table M.26. Operational Rater Agreement Statistics—CSLA Grade 4

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR_1_WE	4	118	87.0	13.0	0.0	0.93	0.02	0.94
PCR_1_WKL	3	118	90.7	9.3	0.0	0.94	0.04	0.94
PCR_2_WE	3	118	92.0	8.0	0.0	0.96	0.01	0.96
PCR_2_WKL	3	118	85.7	14.3	0.0	0.90	0.02	0.90

Table M.27. Field Test Rater Agreement Statistics—CSLA Grade 3

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR_1_WE	3	400	97.8	2.2	0.0	0.98	0.00	0.98
PCR_1_WKL	3	400	91.6	8.4	0.0	0.94	0.02	0.94
PCR_2_WE	3	397	92.1	7.9	0.0	0.93	0.00	0.93
PCR_2_WKL	3	397	94.2	5.8	0.0	0.96	0.02	0.96
PCR_3_WE	3	395	92.1	7.9	0.0	0.94	0.00	0.94
PCR_3_WKL	3	395	90.5	9.5	0.0	0.94	0.00	0.94
PCR_4_WE	3	383	91.3	8.7	0.0	0.93	0.03	0.93
PCR_4_WKL	3	383	91.8	8.2	0.0	0.94	0.01	0.94

Table M.28. Field Test Rater Agreement Statistics—CSLA Grade 4

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR_1_WE	3	584	91.9	8.1	0.0	0.96	0.00	0.96
PCR_1_WKL	3	584	88.4	11.6	0.0	0.94	0.04	0.94
PCR_2_WE	3	587	91.0	9.0	0.0	0.95	0.02	0.95
PCR_2_WKL	3	587	91.4	8.6	0.0	0.95	0.00	0.95

Table M.29. Operational Rater Agreement Statistics—Science Grade 5

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	2	5,519	90.0	9.0	1.0	0.90	0.00	0.90
Item 2	2	5,526	89.7	9.7	0.6	0.86	0.00	0.86
Item 3	2	5,309	91.3	8.0	0.7	0.88	0.00	0.88
Item 4	2	5,307	90.2	9.3	0.5	0.91	0.01	0.91
Item 5	2	5,311	91.1	8.7	0.2	0.93	0.00	0.93
Item 6	2	5,307	89.7	8.8	1.4	0.91	0.00	0.91
Item 7	2	5,531	89.7	10.0	0.4	0.90	0.00	0.90
Item 8	2	5,530	89.9	9.9	0.2	0.90	0.00	0.90
Item 9	2	5,305	90.6	8.4	0.9	0.89	0.00	0.89
Item 10	2	5,303	89.5	10.0	0.5	0.87	0.00	0.87
Item 11	2	5,520	89.8	9.9	0.2	0.83	0.01	0.83
Item 12	2	5,523	90.0	9.8	0.2	0.93	0.00	0.93

Table M.30. Operational Rater Agreement Statistics—Science Grade 8

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	2	4,964	94.4	5.5	0.1	0.92	0.01	0.92
Item 2	2	4,964	92.2	7.7	0.0	0.81	0.00	0.81
Item 3	2	4,782	92.2	7.7	0.1	0.92	0.00	0.92
Item 4	2	4,779	91.1	7.8	1.0	0.91	0.00	0.91
Item 5	2	4,784	91.9	8.0	0.1	0.89	0.00	0.89
Item 6	2	4,785	89.7	10.0	0.3	0.88	0.02	0.88
Item 7	2	4,962	90.0	9.7	0.3	0.93	0.00	0.93
Item 8	2	4,961	89.7	8.9	1.3	0.91	0.01	0.91
Item 9	2	4,962	89.8	9.5	0.6	0.91	0.01	0.91
Item 10	2	4,780	96.8	2.8	0.4	0.95	0.01	0.95
Item 11	2	4,780	92.3	7.5	0.2	0.83	0.00	0.83

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 12	2	4,775	89.6	10.1	0.2	0.87	0.00	0.87
Item 13	2	4,967	89.5	9.6	0.9	0.86	0.01	0.86
Item 14	2	4,966	89.4	9.6	1.0	0.89	0.01	0.89

Table M.31. Operational Rater Agreement Statistics—Science Grade 11

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	2	3,399	92.5	7.4	0.1	0.77	0.01	0.77
Item 2	2	3,398	90.9	8.8	0.3	0.93	0.01	0.93
Item 3	2	3,335	90.7	9.0	0.3	0.93	0.00	0.93
Item 4	2	3,328	89.7	9.9	0.4	0.88	0.00	0.88
Item 5	2	3,404	92.0	7.9	0.2	0.92	0.01	0.92
Item 6	2	3,337	91.0	8.0	1.0	0.92	0.01	0.92
Item 7	2	3,394	90.0	9.5	0.5	0.84	0.01	0.84
Item 8	2	3,399	90.0	9.0	1.0	0.88	0.02	0.88
Item 9	2	3,400	95.2	4.8	0.0	0.86	0.01	0.86
Item 10	2	3,399	89.4	9.5	1.0	0.60	0.01	0.60

Table M.32. Field Test Rater Agreement Statistics—Science Grade 5

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	2	2,999	89.5	10.3	0.2	0.91	0.01	0.91
Item 2	2	3,000	88.2	11.6	0.3	0.90	0.00	0.90
Item 3	2	3,000	89.4	10.0	0.7	0.83	0.01	0.83
Item 4	2	2,999	90.1	8.7	1.2	0.88	0.01	0.88
Item 5	2	2,997	88.5	9.6	1.9	0.88	0.00	0.88
Item 6	2	2,994	88.9	11.0	0.1	0.84	0.02	0.84
Item 7	2	3,000	93.2	6.8	0.0	0.91	0.00	0.91
Item 8	2	3,000	91.4	7.8	0.7	0.93	0.01	0.93
Item 9	2	2,994	96.7	3.3	0.0	0.97	0.01	0.97
Item 10	2	2,998	95.7	4.3	0.0	0.95	0.00	0.95
Item 11	2	2,995	91.6	8.4	0.0	0.94	0.01	0.94
Item 12	2	2,995	92.7	7.3	0.0	0.90	0.00	0.90
Item 13	2	2,995	93.9	5.9	0.2	0.95	0.00	0.95
Item 14	2	2,995	90.0	9.9	0.1	0.91	0.00	0.91
Item 15	2	2,995	89.8	10.0	0.2	0.88	0.01	0.88
Item 16	2	2,996	91.8	8.1	0.1	0.86	0.01	0.86
Item 17	2	2,996	89.1	10.8	0.1	0.82	0.00	0.82
Item 18	2	2,995	94.1	5.8	0.2	0.95	0.00	0.95
Item 19	2	2,994	92.7	7.0	0.3	0.91	0.00	0.91
Item 20	2	2,994	91.4	8.0	0.7	0.90	0.00	0.90
Item 21	2	2,996	94.6	5.1	0.3	0.93	0.01	0.93
Item 22	2	2,996	90.1	9.8	0.0	0.91	0.01	0.91
Item 23	2	2,995	94.0	5.9	0.1	0.93	0.00	0.93

Table M.33. Field Test Rater Agreement Statistics—Science Grade 8

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	2	2,992	90.3	9.3	0.3	0.86	0.01	0.86
Item 2	2	2,994	91.7	8.0	0.3	0.93	0.01	0.93
Item 3	2	2,997	91.0	8.8	0.1	0.93	0.01	0.93
Item 4	2	2,989	95.8	3.3	0.9	0.94	0.00	0.94
Item 5	2	2,996	89.6	10.0	0.4	0.78	0.00	0.78
Item 6	2	2,988	88.9	10.8	0.4	0.79	0.01	0.79
Item 7	2	2,995	89.4	10.5	0.1	0.93	0.00	0.93
Item 8	2	2,993	91.8	7.9	0.2	0.89	0.01	0.89
Item 9	2	2,991	88.6	10.1	1.3	0.71	0.00	0.71
Item 10	2	2,998	86.9	12.4	0.7	0.81	0.01	0.81
Item 11	2	2,986	90.7	9.3	0.1	0.90	0.00	0.90
Item 12	2	2,981	89.9	8.5	1.6	0.78	0.02	0.78
Item 13	2	2,984	93.1	5.7	1.2	0.92	0.01	0.92
Item 14	2	2,987	93.1	6.6	0.3	0.94	0.00	0.94
Item 15	2	2,985	88.9	8.5	2.7	0.89	0.01	0.89
Item 16	2	2,993	87.6	12.2	0.1	0.85	0.02	0.85
Item 17	2	2,983	95.2	4.6	0.2	0.85	0.00	0.85
Item 18	2	2,988	89.4	10.1	0.5	0.90	0.01	0.90
Item 19	2	2,988	89.3	10.4	0.3	0.90	0.00	0.90
Item 20	2	2,982	89.9	9.3	0.8	0.90	0.00	0.90
Item 21	2	2,982	94.2	5.8	0.0	0.94	0.00	0.94

Table M.34. Field Test Rater Agreement Statistics—Science Grade 11

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	2	2,522	73.6	24.4	2.0	0.73	0.03	0.73
Item 2	2	2,493	94.8	5.2	0.0	0.93	0.01	0.93
Item 3	2	2,430	88.2	11.2	0.5	0.90	0.01	0.91
Item 4	2	2,430	90.5	9.0	0.5	0.86	0.01	0.86
Item 5	2	2,484	85.0	14.5	0.5	0.72	0.05	0.72
Item 6	2	2,999	92.4	7.2	0.4	0.89	0.02	0.89
Item 7	2	2,522	89.5	10.5	0.0	0.92	0.01	0.92
Item 8	2	2,341	94.2	5.5	0.3	0.94	0.01	0.94
Item 9	2	2,560	89.5	9.3	1.2	0.78	0.01	0.78
Item 10	2	2,399	91.6	8.2	0.2	0.91	0.00	0.91
Item 11	2	2,471	90.2	8.8	1.0	0.85	0.00	0.85

Appendix N: CMAS Science Grade 11 Blueprint Reduction Study

Colorado Measures of Academic Success (CMAS) Science High School Adjustment Research Report

Pearson

August 31, 2023

Table of Contents

List of Tables and Figures

Section 1. Introduction

The Colorado Department of Education is exploring shortening the Colorado Measures of Academic Success (CMAS) Science High School assessment. The purpose of this research report is to summarize analyses after adjusting the length of the Science assessment in grade 11 for the spring 2023 administration.

A subset of operational items on the spring 2023 assessment was treated as omitted or not administered. Scoring tables or conversion tables were generated based on the adjusted number of operational items. Adjusted scale scores were computed based on student response strings and omitting the selected items. Analyses compared the students' spring 2023 scale scores and performance levels based on the full Science assessment to the adjusted scale scores and performance levels based on the adjusted Science assessment.

Section 2. Methods

This section discusses the data used for the analyses, the adjustments to the Science test form, the item response theory model, generating the scoring tables, and the reported scales and performance levels.

Student Data

The data for this report was spring 2023 Science assessment results in grade 11 for Colorado students. This administration consisted of one computer-based (CBT) operational form and one paper-based (PBT) operational form in addition to several accommodated forms. This study only included the CBT operational form.

The only student records included in this analysis were those used for equating, a total number of 30,738.

Science Assessment Adjustment

Items for the adjusted blueprint were selected to retain the proportion of items from each of the three science standards (Physical, Life, and Earth and Space) as well as the proportion of items that are part of the Science and Engineering Practices (SEP). Table 2.1 provides the number of score points by standard for the full spring 2023 High School Science assessment and the adjusted High School Science assessment. The adjusted test form was reduced by 8 points (from 50 points to 42 points).

Table 2.1. *Science High School Blueprint and Adjusted Points*

Standard	Blueprint Points	Adjusted Points	Blueprint Percentage	Adjusted Percentage
Physical Science	18	15	36%	36%
Life Science	16	14	32%	33%
Earth and Space Science	16	13	32%	31%
Science and Engineering Practices	32-38	29	64%-78%	69%
Total	50	42	100%	84%

Item Response Theory Model

The spring 2023 Science assessment was post-equated. The item parameter estimates from the post-equating analyses were used for both the full Science assessment and the adjusted Science assessment when creating the scoring tables. The operational IRT analyses were conducted by Pearson. The operational items in the incomplete data matrix (IDM) were concurrently calibrated with the two-parameter logistic/three-parameter logistic/generalized partial credit model (2PL/3PL/GPC: Muraki, 1992). The 2PL/GPC is denoted

$$P_{im}(\theta_j) = \frac{\exp\left[\sum_{k=0}^{m} Da_i(\theta_j - b_i + d_{ik})\right]}{\sum_{v=0}^{M_i-1}\exp\left[\sum_{k=0}^{v} Da_i(\theta_j - b_j + d_{iv})\right]}$$

where $a_i(\theta_j - b_i + d_{ik}) = 0$; $P_{im}(\theta_j)$ is the probability of a test taker with θ_j getting score m on item i; M_i is the number of score categories of item i with possible item scores as consecutive integers from 0 to $M_i - 1$; D is the IRT scale constant (1.7). Items calibrated under the 3PL also used a lower-asymptote parameter to account for guessing.

Scoring Tables

IRT ability estimates (θs) are calculated using estimates of item parameters and thetas are substituted for parameters in the formulas in the generalized partial credit model for both dichotomous and polytomous items. These estimates are then linearly transformed to the reporting scale using a slope of 24.6161 and an intercept of 730.79 which were determined by relating two of the established cut scores on the theta metric to their fixed values on the reporting scale. The estimated conditional standard error of measurement (CSEM) for each scale score is computed.

All operational procedures for generating the conversion files were followed to generate the scale scores for the adjusted assessment.

Reporting Scales and Performance Levels

The CMAS Science reporting scales designate student performance into one of four Performance Levels that delineate the knowledge, skills, and practices students are able to demonstrate. Level 1 indicates the lowest level of performance and Level 4 indicates the highest level of performance:

- Level 1: Partially met expectations
- Level 2: Approached expectations

- Level 3: Met expectations
- Level 4: Exceeded expectations

Summative scale scores, which reflects performance across all items on the assessment, range from 650 to 850 and categorize students into one of four summative performance levels with a 725 representing the threshold of Level 2, and 750 representing the threshold of Level 3 which represents college and career readiness (CCR). The threshold score for Level 4 is 787 for the High School Science assessment.

Science has a subset of skills, or standards, in which additional information regarding student performance is provided. The standards performance levels categorize students into one of three levels based on the average performance of students at the summative Performance Level 3 and Level 4: *Lower than Average*, *Average* or *Higher than Average*. The standards performance levels provide information regarding targeted instructional needs.

Section 3. Results

This section presents the results for the Science adjusted assessment in comparison to the full spring 2023 administration. The results include scale score summary statistics, overall performance level agreement, standard performance level agreement, correlations, overall and standards test characteristic curves.

Scale Score Summary Statistics

The overall adjusted scale score and adjusted scale score conditional standard errors of measurement (CSEMs) were calculated based on all operational items. Table 3.1 reports summary statistics (count, mean, standard deviation, minimum, and maximum) for the full and adjusted scale scores and CSEM.

The average scale scores were similar for the adjusted and full test forms. The average scale score difference was less than 0.70, the average conditional standard errors were slightly lower for the longer test form for the overall scale score. The average conditional standard errors for the physical science standard and SEP scale scores were lower for the longer test and quite a bit lower for the earth and space science and life science standards.

Table 3.1. *Summary Statistics for Full and Adjusted Scores for CMAS Science High School*

		Count	Mean	Standard Deviation	Minimum	Maximum
Physical Science	Full Scale Score	30,738	476.68	34.29	400	550
	Adjusted Scale Score	30,738	476.61	34.88	400	550
	Full Scale Score CSEM	30,738	33.19	61.94	9	224
	Adjusted Scale Score CSEM	30,738	35.16	62.85	11	225
Life Science	Full Scale Score	30,738	476.41	35.80	400	550
	Adjusted Scale Score	30,738	476.02	37.04	400	550
	Full Scale Score CSEM	30,738	46.95	93.28	11	314
	Adjusted Scale Score CSEM	30,738	68.75	141.29	12	443
Earth and Space Science	Full Scale Score	30,738	475.60	35.98	400	550
	Adjusted Scale Score	30,738	475.00	37.17	400	550
	Full Scale Score CSEM	30,738	36.73	60.91	11	205
	Adjusted Scale Score CSEM	30,738	46.58	77.28	12	241
SEPs	Full Scale Score	30,738	477.88	31.16	400	550
	Adjusted Scale Score	30,738	477.66	31.82	400	550
	Full Scale Score CSEM	30,738	18.42	36.00	7	175
	Adjusted Scale Score CSEM	30,738	24.05	49.74	8	219
Overall	**Full Scale Score**	**30,738**	**728.81**	**29.43**	**650**	**842**
	Adjusted Scale Score	**30,738**	**728.79**	**29.69**	**650**	**840**
	Full Scale Score CSEM	**30,738**	**9.97**	**9.00**	**6**	**47***
	Adjusted Scale Score CSEM	**30,738**	**10.91**	**9.18**	**7**	**47***

**The maximum CSEM for the overall scale score was limited to 47 following operational procedure for Spring 2023*

Performance Level Agreement

Table 3.2 lists the percent of students assigned the exact same performance level for both the full and the adjusted Science assessment by standard. In addition, Table 3.2 lists the percent of students assigned to different performance levels between the full and adjusted Science assessment overall and across standards. If the adjusted performance levels were a higher ability level compared to the full performance level the number and percent of students are listed as “Higher Level for Adjusted”. If the adjusted performance levels were a lower ability level compared to the full performance level the number and percent of students are listed as “Lower Level for Adjusted”.

The percent of exact agreement in the overall performance level designation between the full assessment and the adjusted assessment for any standard ranged from 91.5% – 93.5%. The percent of students in the Higher Level or the Lower Level for adjusted for any of the standards ranged from 2.8% – 5.2%.

Table 3.2. *CMAS Science High School Performance Level Agreement by Standard*

Standard	Exact Agreement	Higher Level for Adjusted	Lower Level for Adjusted
Physical Science	93.5%	3.7%	2.8%
Life Science	91.5%	3.3%	5.2%
Earth and Space Science	93.1%	3.1%	3.8%
Science and Engineering Practices	92.6%	4.2%	3.2%
Overall	**93.1%**	**3.5%**	**3.4%**

Tables 3.3 – 3.6 show the number and percent of students by the full performance level designation and the adjusted performance level designation for the overall test and each standard. The values bolded in the tables represent exact agreement. For all standards and the overall test, if the performance level designation was not exact, the difference was always within an adjacent performance level.

Table 3.3. *CMAS Science High School Overall Performance Level Percent Agreement*

Overall	Adjusted L1	Adjusted L2	Adjusted L3	Adjusted L4	**Total**
2023 L1	**11,460 (37.3%)**	519 (1.7%)			11,979 (39.0%)
2023 L2	585 (1.9%)	**10,216 (33.2%)**	490 (1.6%)		11,291 (36.7%)
2023 L3		427 (1.4%)	**6,750 (22.0%)**	69 (0.2%)	7,246 (23.6%)
2023 L4			29 (0.1%)	**193 (0.6%)**	222 (0.7%)
Total	12,045 (39.2%)	11,162 (36.3%)	7,269 (23.7%)	262 (0.9%)	**30,738 (100%)**

Table 3.4. *CMAS Science High School Physical Science Performance Level Percent Agreement*

Physical Science	Adjusted L1	Adjusted L2	Adjusted L3	**Total**
2023 L1	**3,968 (12.9%)**	440 (1.4%)		4,408 (14.3%)
2023 L2	333 (1.1%)	**21,126 (68.7%)**	694 (2.3%)	22,153 (72.1%)
2023 L3		529 (1.7%)	**3,648 (11.9%)**	4,177 (13.6%)
Total	4,301 (14.0%)	22,095 (71.9%)	4,342 (14.1%)	**30,738 (100%)**

Table 3.5. *CMAS Science High School Life Science Performance Level Percent Agreement*

Life Science	Adjusted L1	Adjusted L2	Adjusted L3	**Total**
2023 L1	**4,068 (13.2%)**	501 (1.6%)		4,569 (14.9%)
2023 L2	903 (2.9%)	**20,439 (66.5%)**	506 (1.7%)	21,848 (71.1%)
2023 L3		697 (2.3%)	**3,624 (11.8%)**	4,321 (14.1%)
Total	4,971 (16.2%)	21,637 (70.4%)	4,130 (13.4%)	**30,738 (100%)**

Table 3.6. *CMAS Science High School Earth and Space Science Performance Level Percent Agreement*

Earth and Space Science	Adjusted L1	Adjusted L2	Adjusted L3	**Total**
2023 L1	**3,975 (12.9%)**	291 (1.0%)		4,266 (13.9%)
2023 L2	713 (2.3%)	**21,067 (68.5%)**	677 (2.2%)	22,457 (73.1%)
2023 L3		454 (1.5%)	**3,561 (11.6%)**	4,015 (13.1%)
Total	4,688 (15.3%)	21,812 (71.0%)	4,238 (13.8%)	**30,738 (100%)**

Table 3.7. *CMAS Science High School Science and Engineering Practices Performance Level Percent Agreement*

SEPs	Adjusted L1	Adjusted L2	Adjusted L3	**Total**
2023 L1	**3,556 (11.6%)**	716 (2.3%)		4,272 (13.9%)
2023 L2	543 (1.8%)	**20,947 (68.2%)**	578 (1.9%)	22,068 (71.8%)
2023 L3		438 (1.4%)	**3,960 (12.9%)**	4,398 (14.3%)
Total	4,099 (13.3%)	22,101 (71.5%)	4,538 (14.8%)	**30,738 (100%)**

Correlations and Scale Score Differences

The correlation between the full scale scores and adjusted scale scores for the overall test and each standard. Correlations ranged between .951 and .986.

Table 3.8. *Pearson Correlations between Full and Adjusted Scale Scores*

Standard	Full and Adjusted Scale Scores Correlation
Physical Science	0.983
Life Science	0.951
Earth and Space Science	0.964
Science Engineering and Practices	0.974
Overall	**0.986**

Figures 3.1 – 3.10 display the scatterplot of full scale scores versus adjusted scale scores and the frequency distribution of the differences in scale scores between full and adjusted.

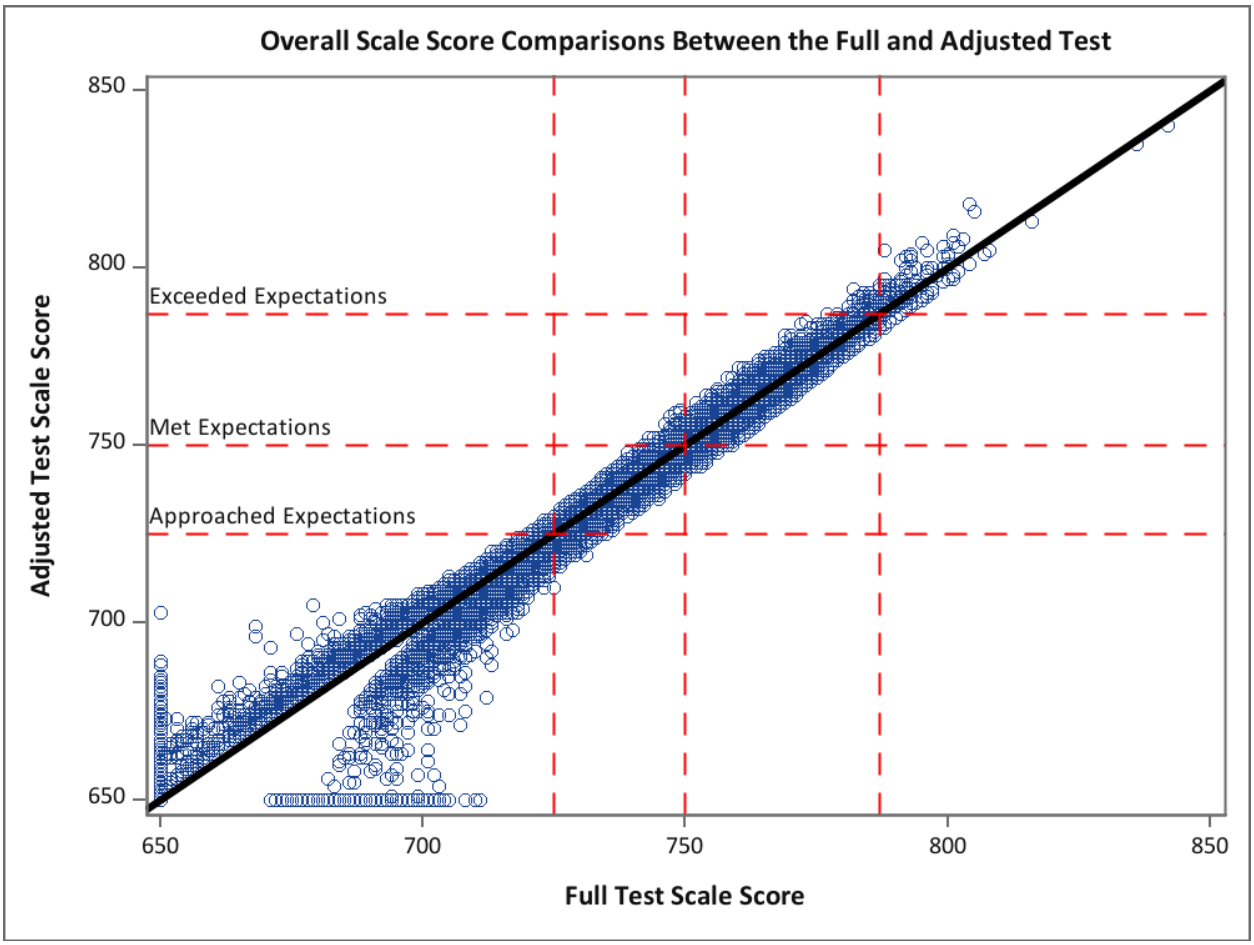


Figure 3.1. CMAS Science High School Overall Full vs. Adjusted Scale Scores.

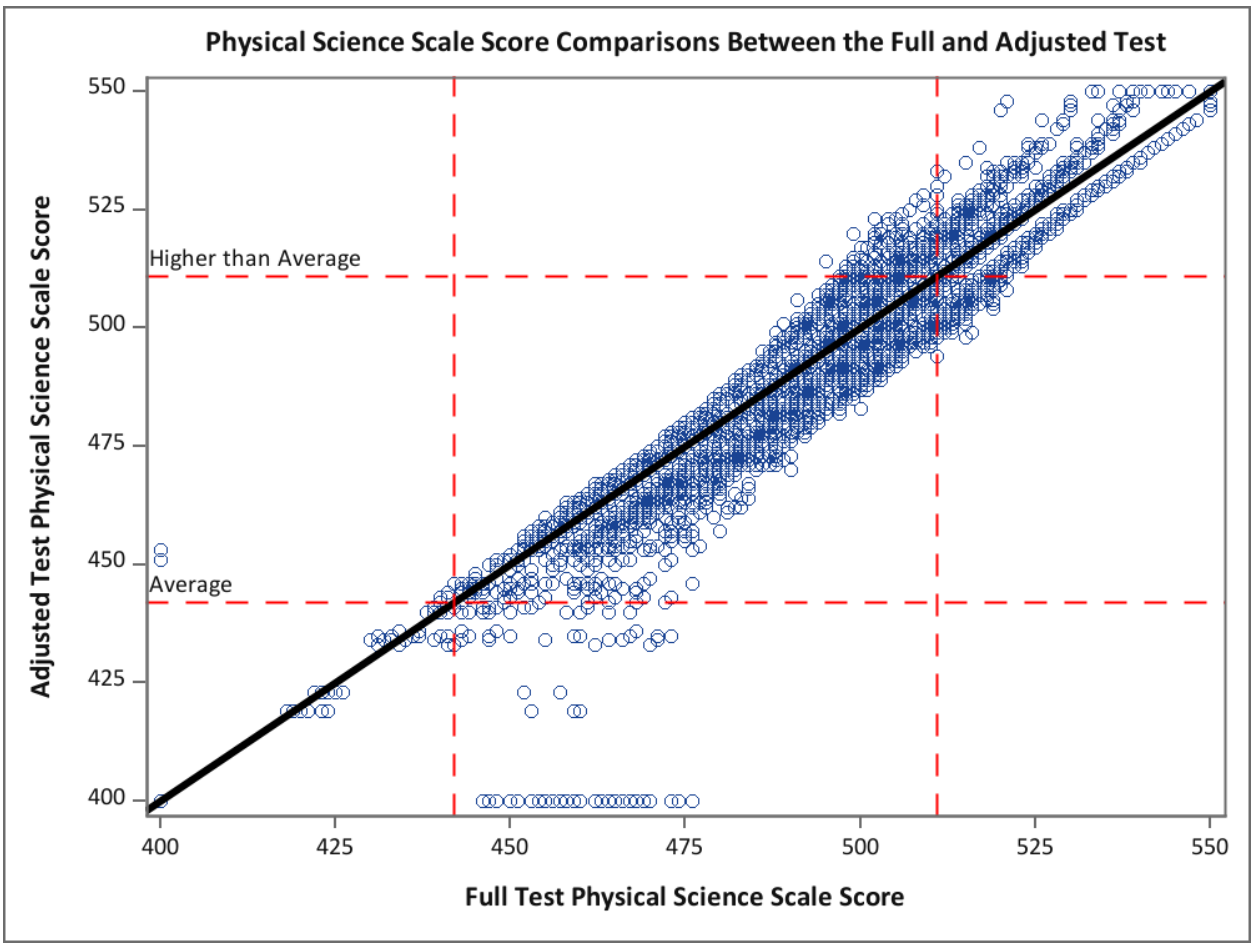


Figure 3.2. CMAS Science High School Physical Science Full vs. Adjusted Scale Scores.

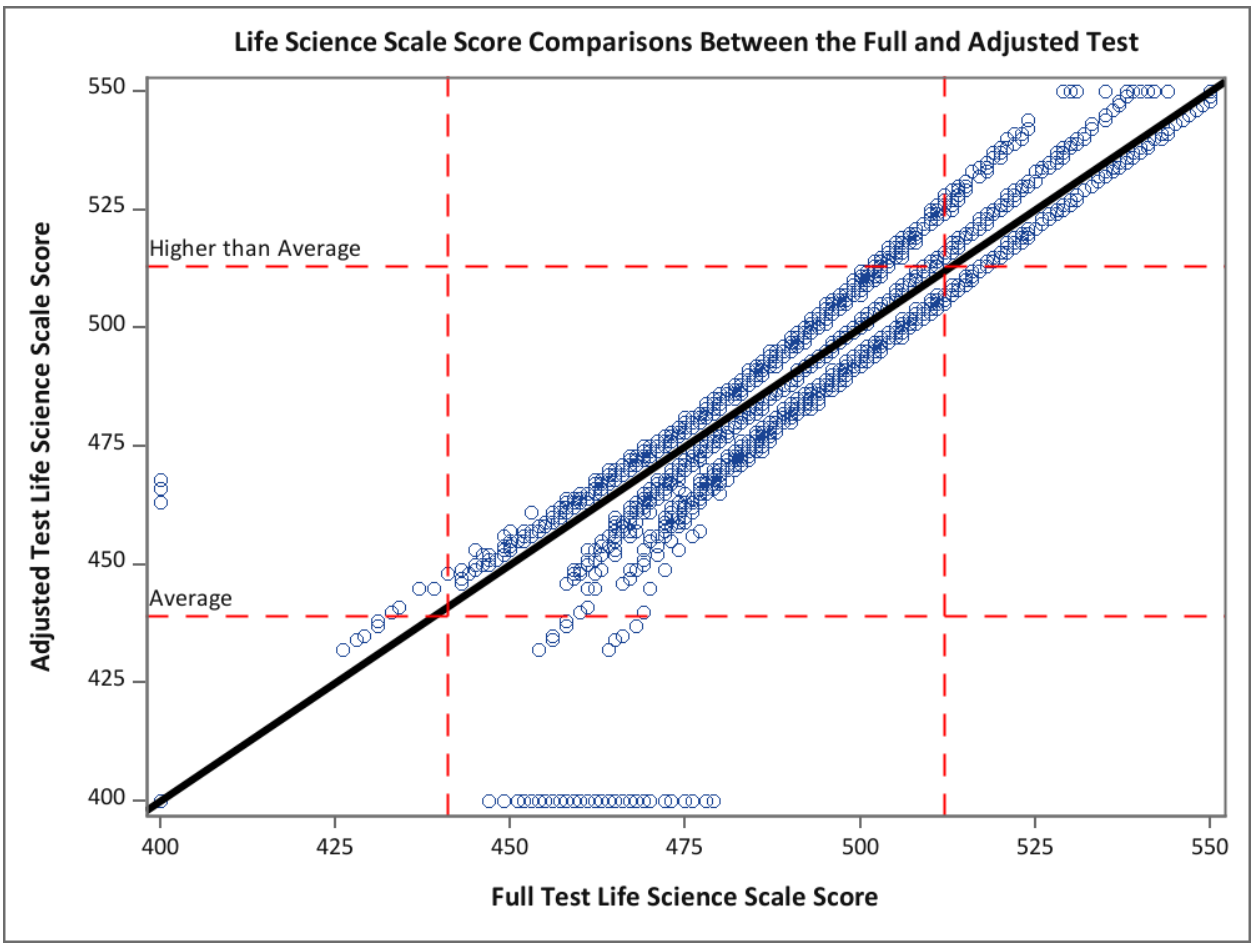


Figure 3.3. CMAS Science High School Life Science Full vs. Adjusted Scale Scores.

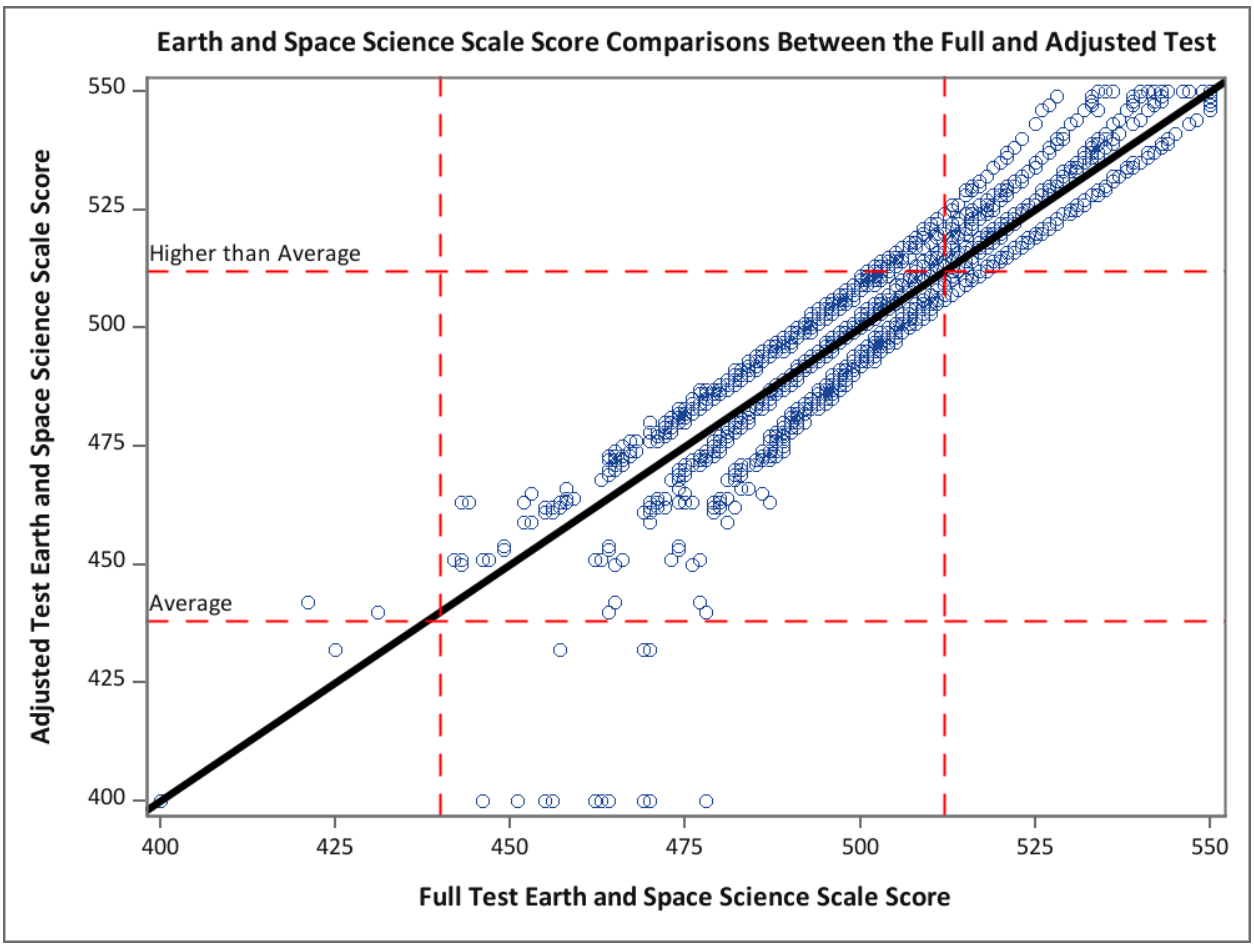


Figure 3.4. CMAS Science High School Earth and Space Science Full vs. Adjusted Scale Scores.

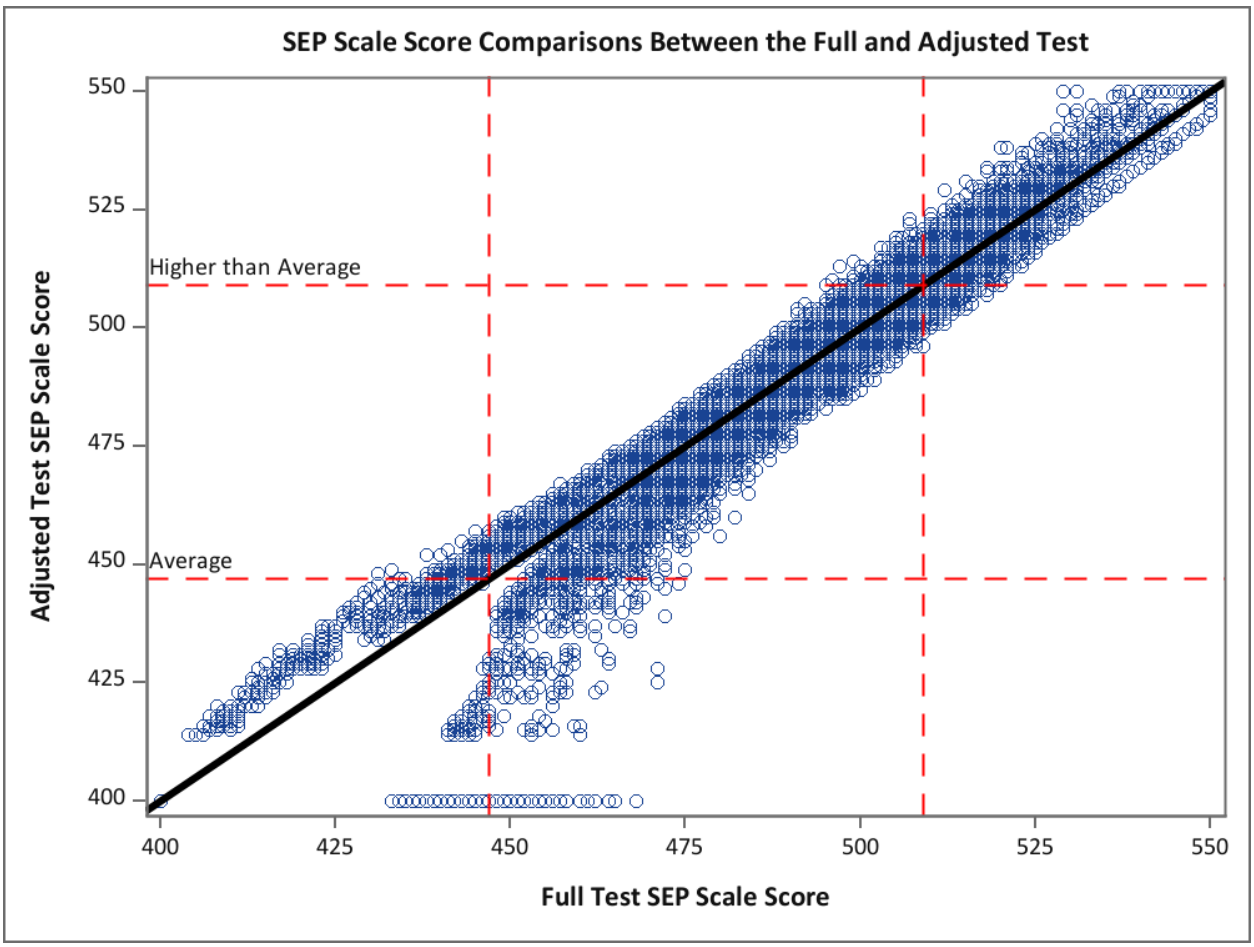


Figure 3.5. CMAS Science High School Science and Engineering Practices Full vs. Adjusted Scale Scores

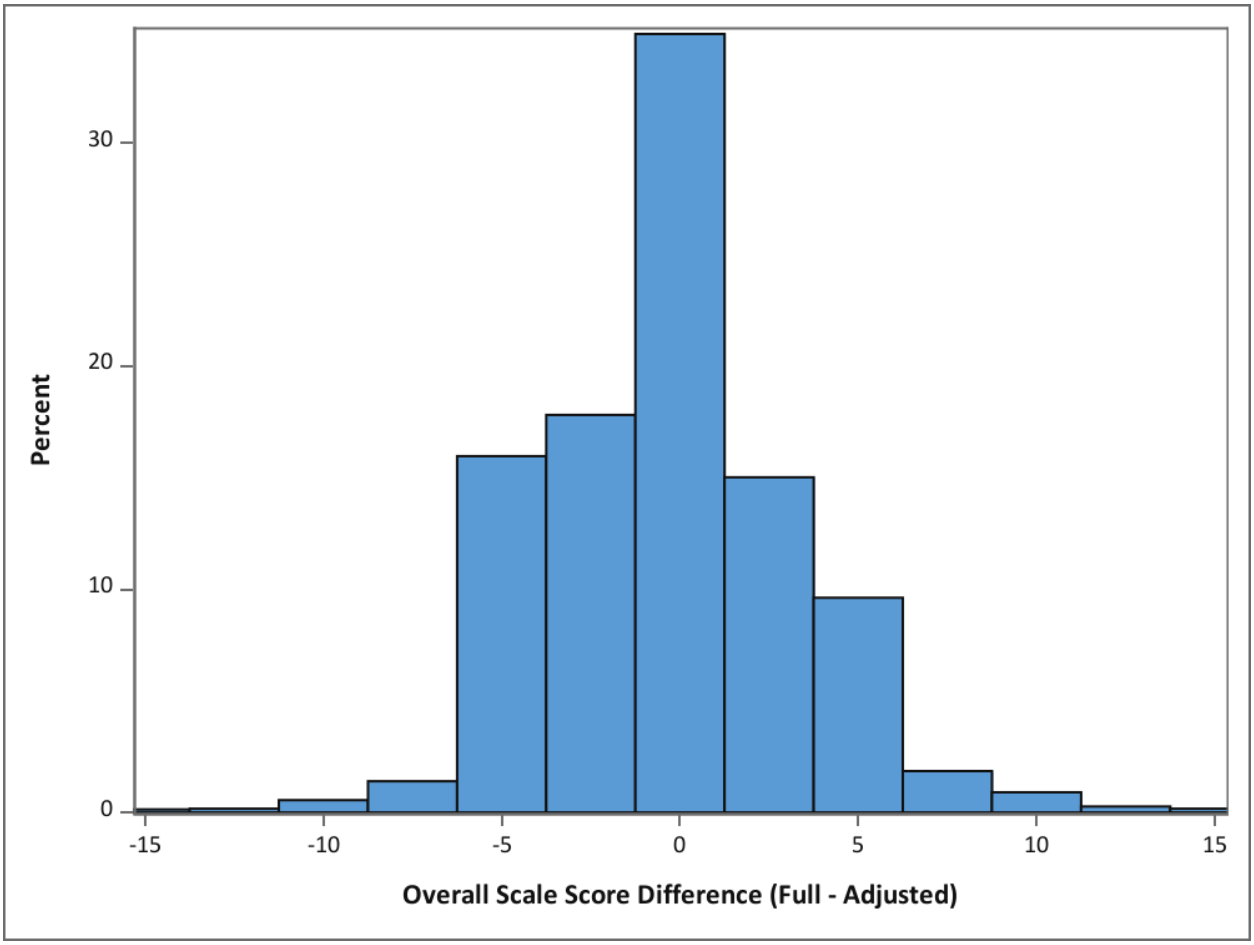


Figure 3.6. CMAS Science High School Overall Scale Score Differences Between Adjusted and Full Scale Scores.

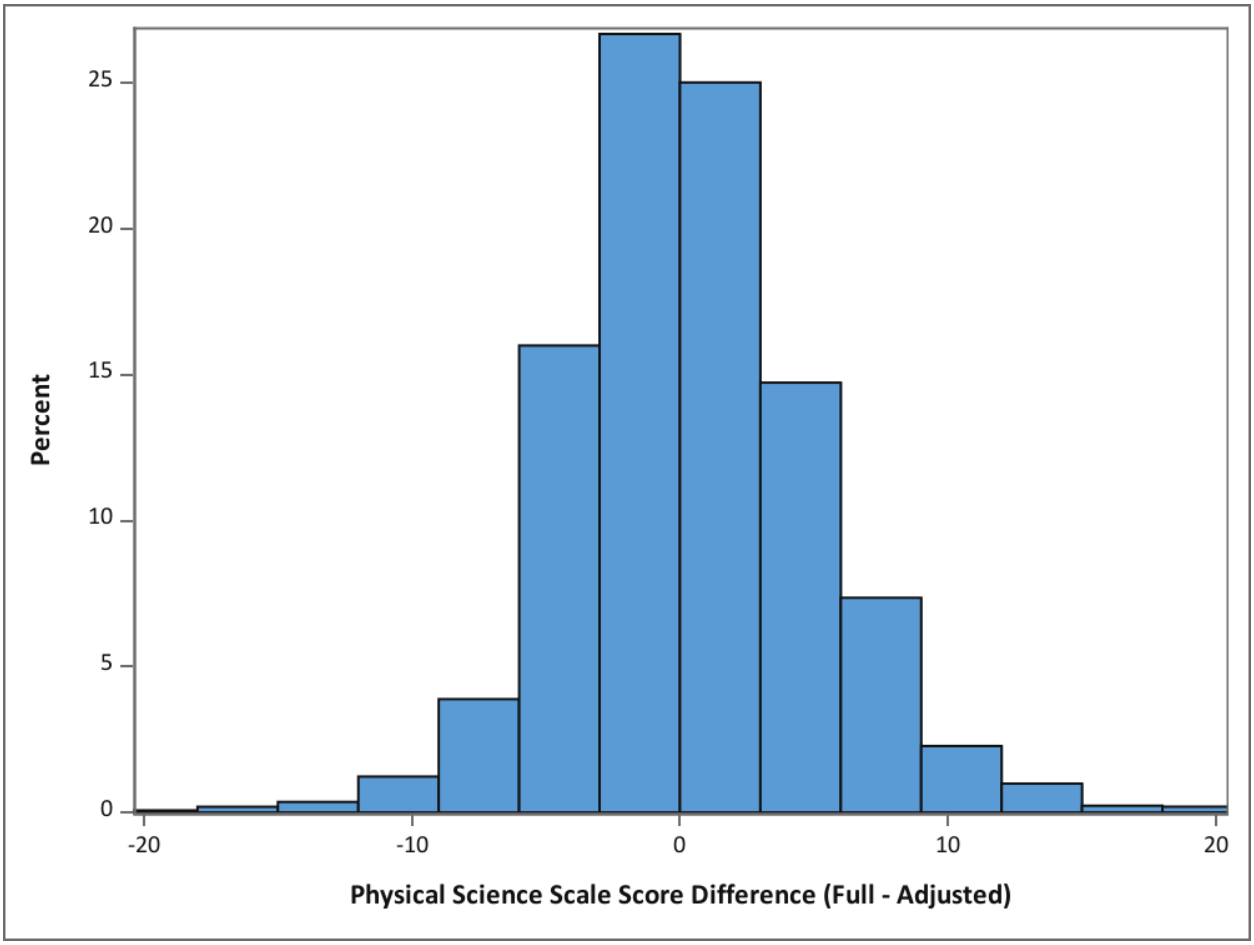


Figure 3.7. CMAS Science High School Physical Science Scale Score Differences Between Adjusted and Full Scale Scores.

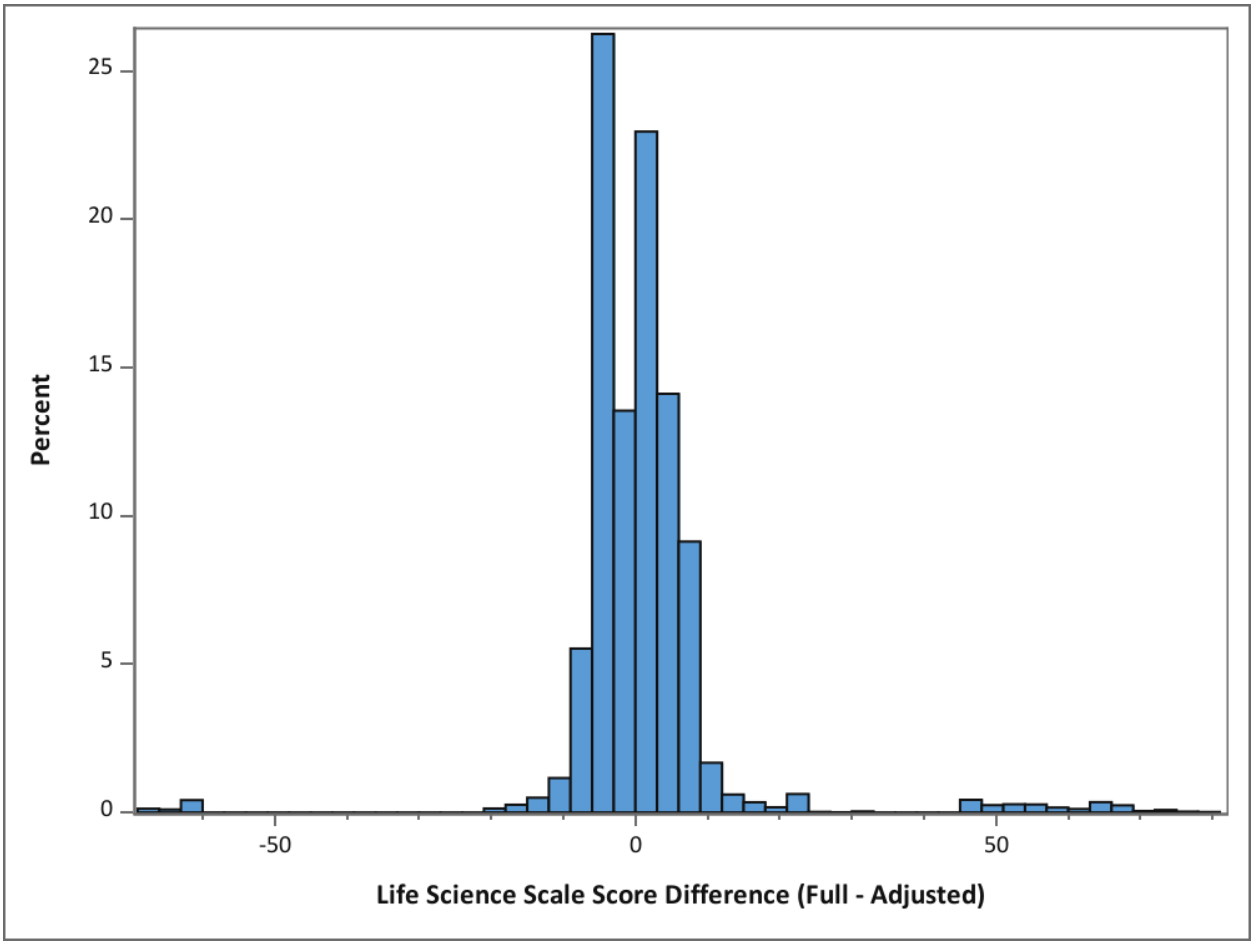


Figure 3.8. CMAS Science High School Life Science Scale Score Differences Between Adjusted and Full Scale Scores.

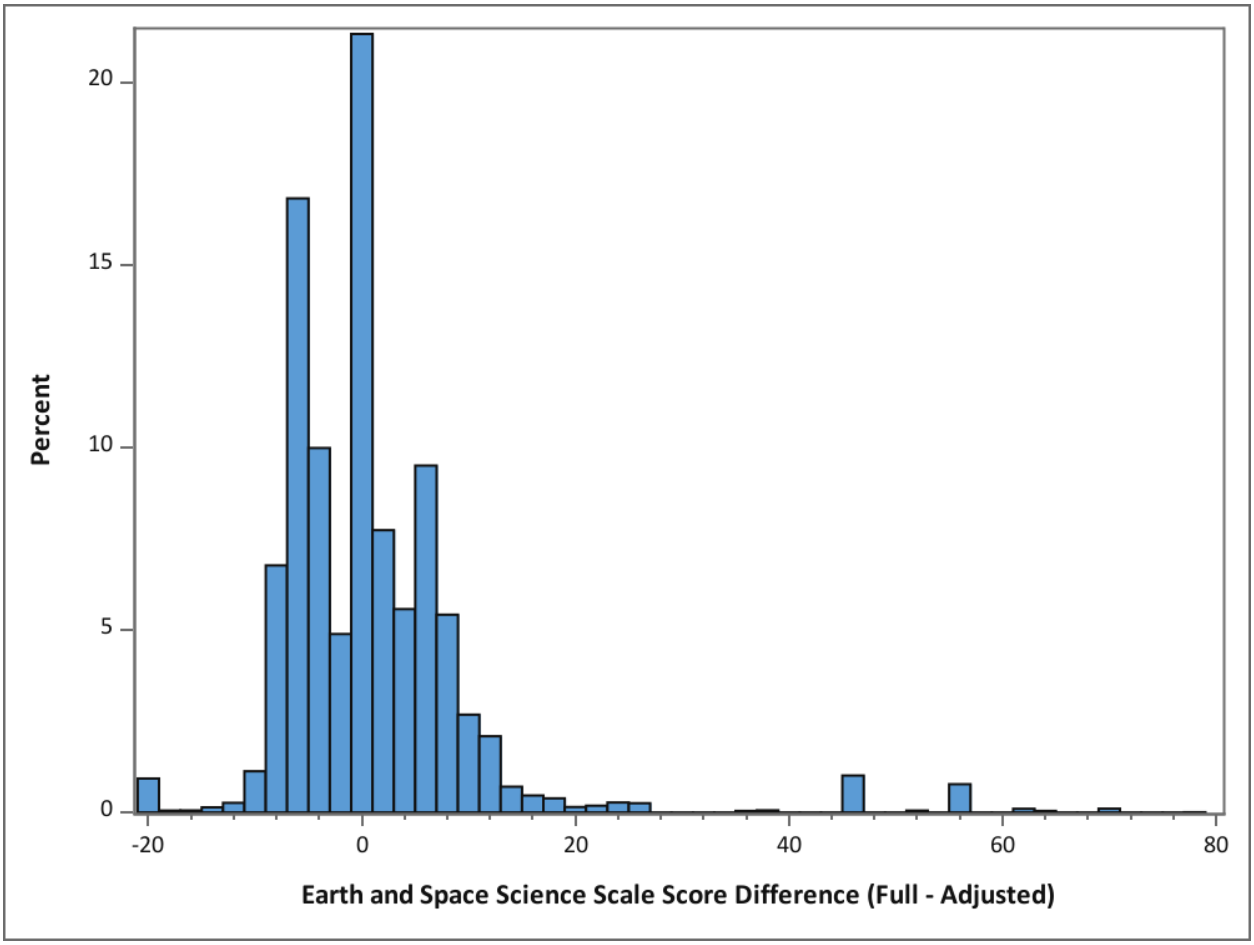


Figure 3.9. CMAS Science High School Earth and Space Science Scale Score Differences Between Adjusted and Full Scale Scores.

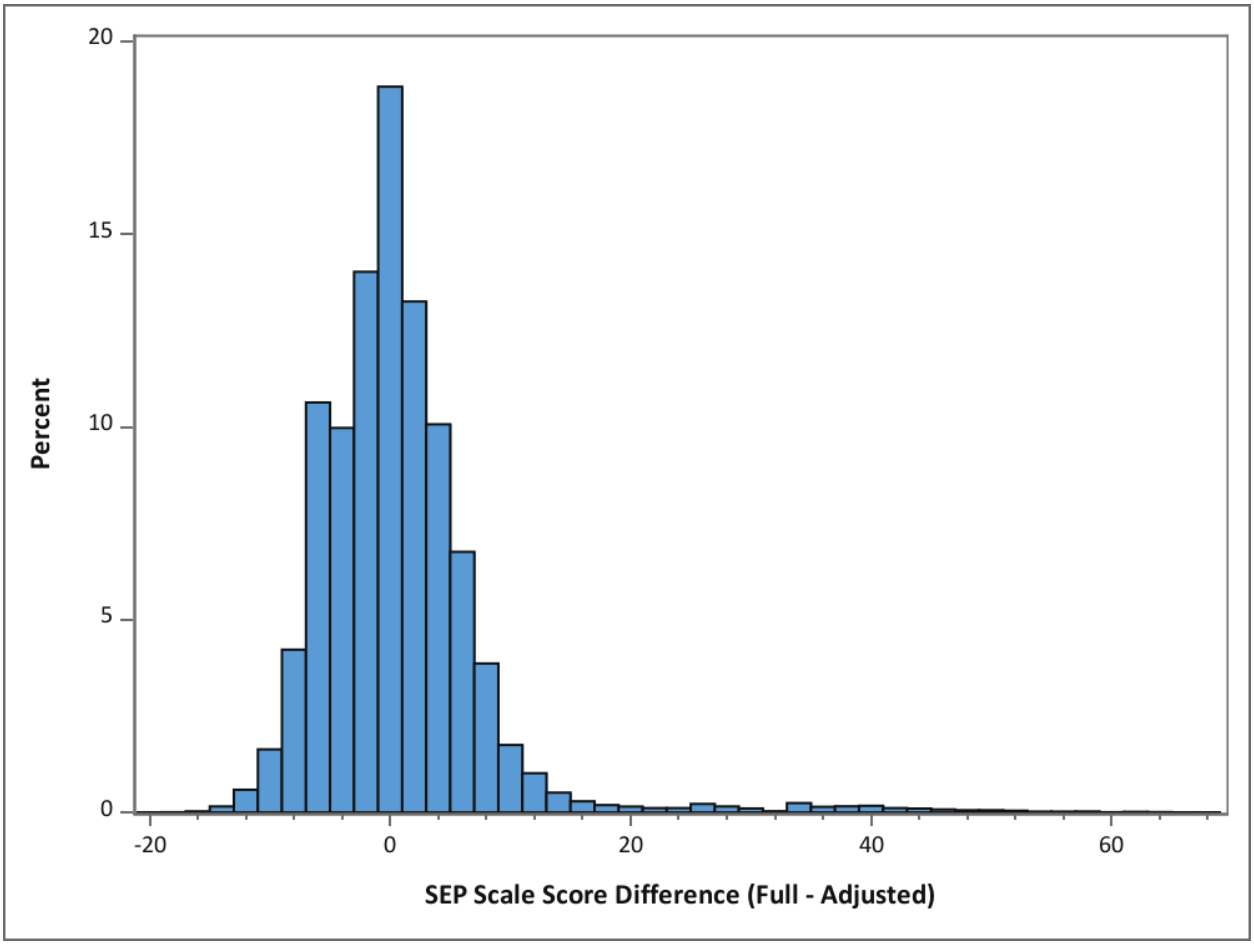


Figure 3.10. CMAS Science High School Science and Engineering Practices Scale Score Differences Between Adjusted and Full Scale Scores.

Summative Test Characteristic Curves

The test characteristic curves were generated for the raw score to theta scale for the overall theta scale and for each standard. Due to the difference in raw score total and the number of operational items across the full and adjusted forms, TCCs are provided based on the percent of the total maximum possible score points. Figures 3.11 – 3.15 present the test characteristic curves for the overall test and each standard. In general, the test characteristic curves for the adjusted test forms are similar to the full test form test characteristic curves. Only one set of curves resulted in differences that were greater than 5%. For Physical Science the difference was between 5.03% and 5.43% where theta is between 2.4 and 5.

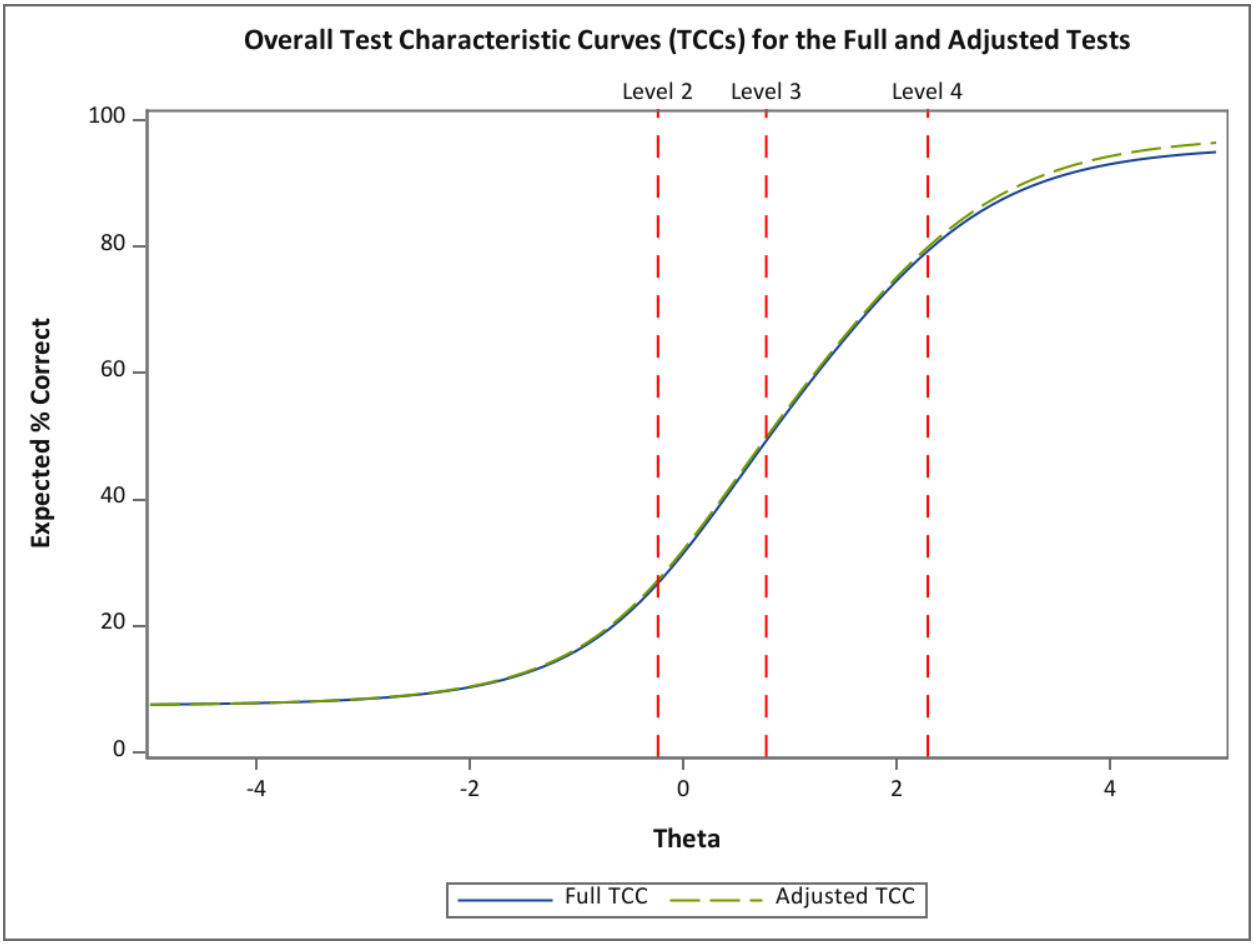


Figure 3.11. CMAS Science High School Overall Percentage Test Characteristic Curves for Full and Adjusted Raw Scores.

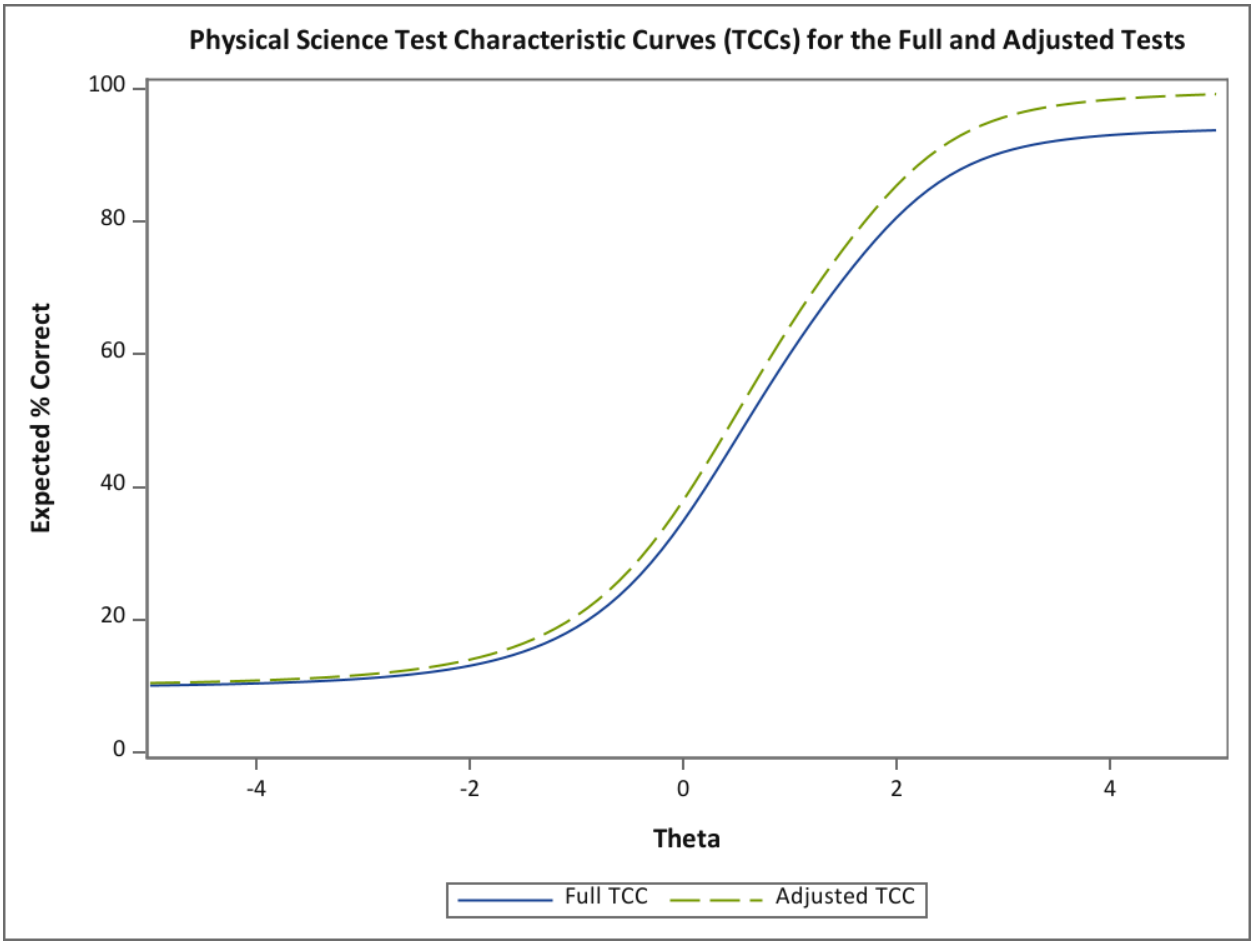


Figure 3.12. CMAS Science High School Physical Science Percentage Test Characteristic Curves for Full and Adjusted Raw Scores.

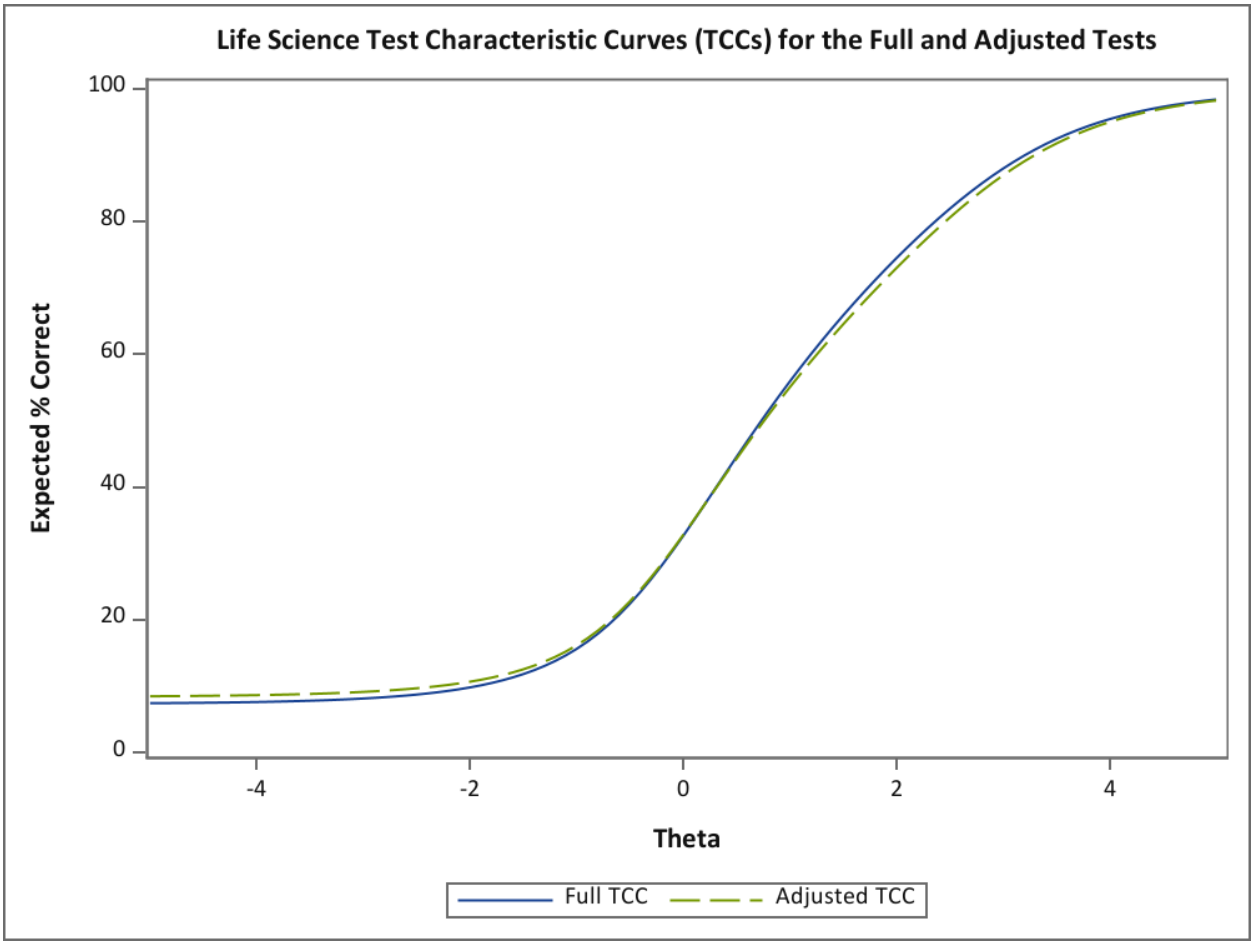


Figure 3.13. CMAS Science High School Life Science Percentage Test Characteristic Curves for Full and Adjusted Raw Scores.

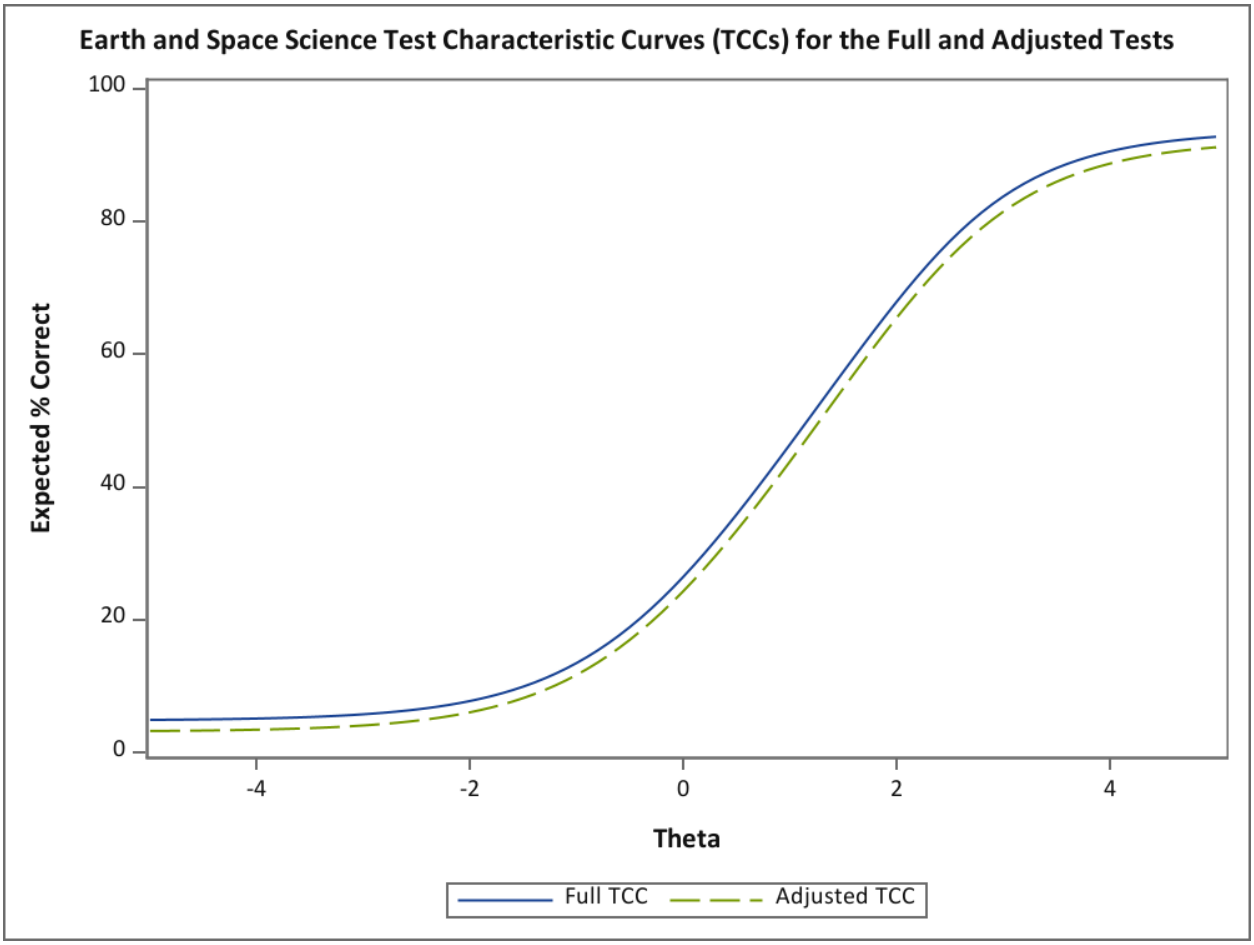


Figure 3.14. CMAS Science High School Earth and Space Science Percentage Test Characteristic Curves for Full and Adjusted Raw Scores.

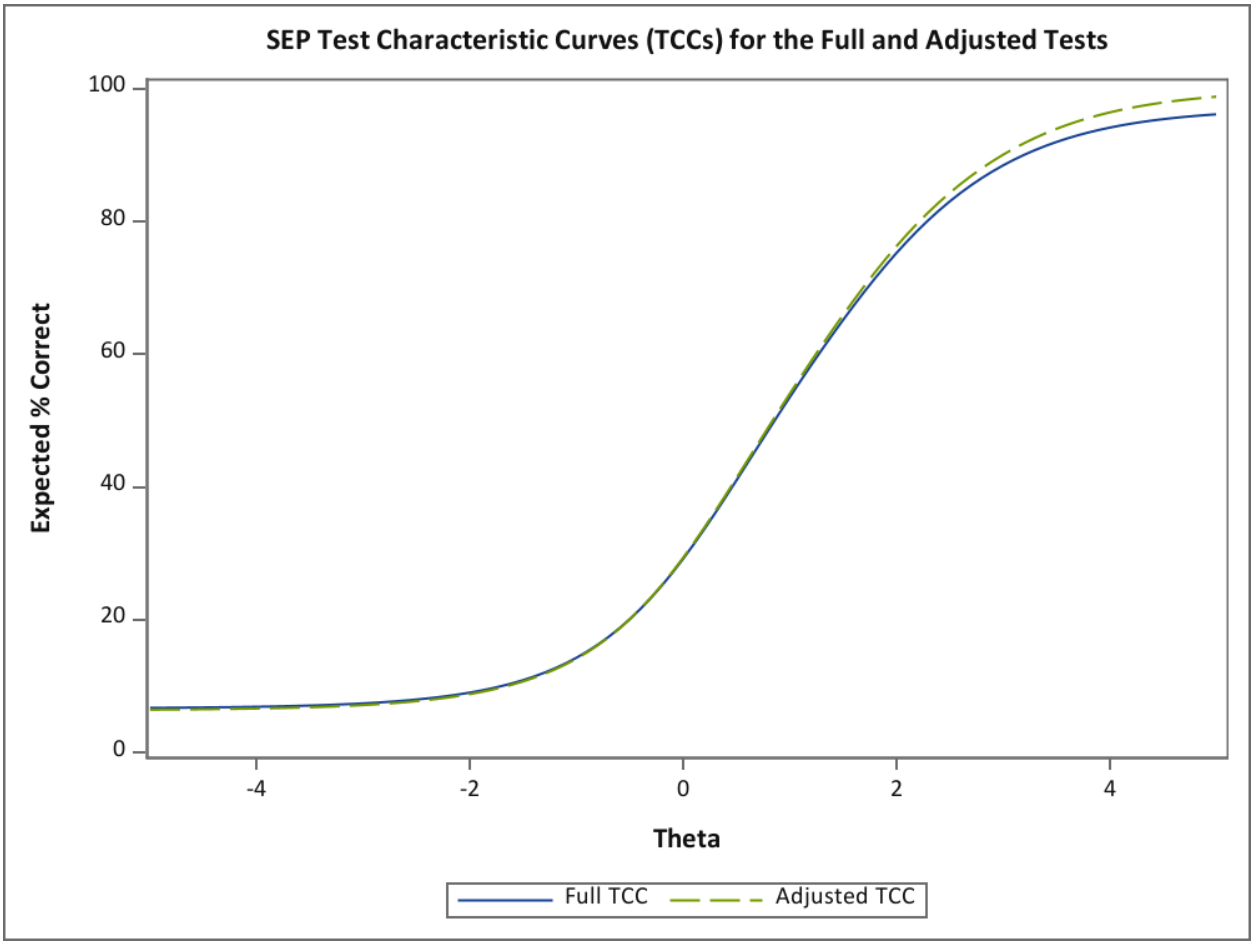


Figure 3.15. CMAS Science High School Science and Engineering Practices Percentage Test Characteristic Curves for Full and Adjusted Raw Scores.

Section 4. Summary

This analysis used the spring 2023 administration of the grade 11 CMAS Science assessment for Colorado students to study the potential impact of omitting some items in future administrations in order to reduce testing time. 8 points were removed proportionally by standard. Analyses compared the students' spring 2023 scale scores and performance levels based on the full Science assessment to the adjusted scale scores and performance levels based on the adjusted Science assessment.

The average scale scores were similar for the adjusted and full test forms. The percent of exact agreement in the overall performance level designation between the full assessment and the adjusted assessment ranged from 91.5% – 93.5% exact agreement across the standards and overall test. In addition, the correlations were all greater than .95.

A potential limitation to this study is that the items that could be omitted were constrained to those in the 2023 administration. In addition, having fewer score points may impact the precision of performance level classification.

References

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16,* 159–176.

Appendix O: CMAS Science 2024 Cognitive Lab Study

Analysis of the Colorado Science Assessment Cognitive Interviews

Prepared by Dr. Muireann Hendriksen

Learning Impact Measurement, Efficacy & Learning

March 2024

Background

The 2020 Colorado Academic Standards (CAS) in science represent what all Colorado students should know and be able to do in science as a result of their preschool through twelfth-grade science education. Specific expectations are given for students who complete each grade from preschool through eighth grade and for high school. These standards outline the essential level of science content knowledge and the application of the skills needed by all Colorado citizens to participate productively in our increasingly global, information-driven society.

To inform this process, the revision committee leveraged numerous resources, including "A Framework for K-12 Science Education", released by the National Academies of Science in 2012. This publication synthesizes over 20 years of science education research and established the context for three-dimensional science standards that emphasizes a student-centered approach to science teaching and learning that integrates the practices, core ideas, and crosscutting concepts to support a high-quality science education that is inclusive and accessible for all Colorado students. A Framework for K-12 Science Education indicates the inextricable ties between science and everyday phenomenon,

> *"Anchoring learning in explaining phenomena supports student agency for wanting to build science and engineering knowledge. Students are able to identify an answer to "why do I need to learn this?" before they even know what the "this" is. In contrast, students might not understand the importance of learning science ideas that teachers and curriculum designers know are important but that are unconnected from phenomena. By centering science education on phenomena that students are motivated to explain, the focus of learning shifts from learning about a topic to figuring out why or how something happens. For example, instead of simply learning about the topics of photosynthesis and mitosis, students are engaged in building evidence-based explanatory ideas that help them figure out how a tree grows."* - Using Phenomena in NGSS-Designed Lessons and Units.

The Colorado Measures of Academic Success (CMAS) science is Colorado's standard-based assessment designed to measure the Colorado science standards. The CMAS incorporates phenomena in three ways:

- Simulation Clusters: Students are presented with an interactive simulation of a science model or experiment and asked to manipulate the simulation to make sense of the phenomenon shown and answer multiple associated two- or three-dimensional questions using their knowledge of the 2020 CAS.
- Static clusters: Students are presented with background information, models, images, graphs, tables, and additional media and asked to engage with the material to make sense of the phenomenon described and answer multiple associated two- or three-dimensional questions using their knowledge of the 2020 CAS.
- Standalone Items: Students are presented with a unique phenomenon and asked to make sense of that phenomenon based on the information in the stimulus and answer the two- or three-dimensional question using their knowledge of the 2020 CAS.

Given, then, the importance of phenomena in delivering science education, it is of critical importance to understand students' engagement with phenomenon-based assessment materials, both in terms of degree of engagement and the methods and cognitive processes by which students approach that engagement. A strong understanding of how students engage with assessment phenomena will further success in both development of stimuli and interpretation of testing outcomes.

Research Questions

This study aims to answer the following questions:

1. To what extent do students need and engage with the directions provided for simulations on the CMAS science assessments?
2. Do the assessment stimuli elicit the intended cognitive response process?
3. Do students find simulative clusters more, less, or equally engaging to static clusters?
4. What is the level of difficulty or ease in navigating the different types of stimulus information to answer the associated questions?
5. Do the accessibility tools (Text-To-Speech, TTS) and features (design of art and layout) produce an inadvertent effect on the construct assessed? i.e. Does the tool help or hamper student's ability to answer the question?

Study Design

The cognitive interviews were conducted virtually by trained content experts from Pearson. Participants completed the interviews during one class period on a regular school day, in a quiet or private room in their school building. A representative from the Colorado Dept. of Education (CDE) was present in the room with participants to facilitate the interview process. The interviews were administered using a private interview link that only the CDE and Pearson representatives had access to. All interviews were recorded via Microsoft Teams to enable analysis of participants' actions on the items (such as where a student clicks, how they move through the items), their non-verbal behavior, and transcription of the interviews.

The cognitive interviews were performed using a sample of one simulation, one static cluster and items aligned to these grade 5 and 8 phenomena. The simulations were chosen to reflect continual improvement of presentation, showcasing the most recent iterations of simulation layout and interactions. In order to understand the extent to which students need and engage with directions, participants were shown the same simulation without directions and were given the directions after completing the items associated with the sims. They were then asked about their experience and preferences regarding the utility and necessity of the directions.

In recognition that students may be completing items above their grade level (e.g. Grade 4 students completing the protocol for Grade 5), stimulus material was also chosen based on ease of interpretation for students at lower grade levels. The items selected were either released or slated for release, so that participants in Grades 4 and 7 at the time of the interviews would not have an unfair advantage when taking the assessment the following year. Participants were also given two warm-up items without a stimulus at the start of the interview, to ensure they understood the think-aloud process.

Students were asked to verbalize their thoughts out loud while reading and responding to the items. Their responses to the items were captured by the test delivery system, and interviewers were encouraged to probe student thinking as needed once a student responded to an item, to gain a better understanding of their thought process. After each interview, interviewers completed a debriefing form to record additional thoughts on the interviews and note any issues or glitches if relevant.

Recruitment

The study recruited students from grades 4 and 6 for the grade 5 protocol, and students from grades 7 and 9 for the grade 8 protocol. The study aimed to recruit a sample representative of the Colorado student population to the extent possible and aimed to recruit 3-4 students from each participating grade and school. This included one student per grade and school who used TTS. Thus, the study aimed to recruit 9-12 students per grade, and 27-36 students in total.

28 students participated in the Cognitive Interviews. The following tables describe the demographic profile of participants.

- 16 students participated in the Grade 5 cognitive interviews:

Regions	• 3 schools Metro area • 1 school North Central • 1 school Pikes Peak
Grades	• 7 students were in Grade 4 at the time of the interview • 9 students were in Grade 6
Past Year Achievement in Math or Science	• 4 Approached • 4 Partially Met • 5 Met • 3 Exceeded
TTS	• 3 students used TTS
Gender	• 9 male students • 7 female students
Race/ Ethnicity	• 8 White • 4 Hispanic • 3 Asian • 1 Multiracial
Socioeconomic Status	• 8 students were from economically disadvantaged backgrounds

- 12 students participated in the Grade 8 cognitive interviews:

Schools	• 1 school North Central • 2 schools Pikes Peak
Grades	• 8 students were in Grade 7 at the time of interview; • 4 students were in Grade 9
Prior Achievement	• 5 Approached • 2 Partially Met • 3 Met • 2 Unknown
TTS	• 3 students used TTS
Gender	• 5 male students • 7 female students
Race/ Ethnicity	• 8 White • 1 Hispanic • 1 Asian • 2 Multiracial
Socioeconomic Status	• 1 student was from an economically disadvantaged background

Analysis

Framework analysis (Ritchie & Spencer, 1994) was chosen as the most appropriate form of analysis, given that it was developed for the explicit purpose of analyzing qualitative data in applied settings. Framework analysis is an inherently comparative form of thematic analysis which employs an organized structure (i.e., a framework) to conduct cross-sectional analysis using a combination of data description and abstraction. This allows the researcher to identify, describe, and interpret key patterns within and across cases relating to the phenomena of interest.

Analysis followed the 5-stage framework outlined by Ritchie & Spencer (1994); familiarization; identifying a framework; indexing; charting; and mapping and interpretation. At the familiarization stage, the researcher read through all interview transcripts to 'immerse' themselves in the data. They also consulted the interviewer debriefing forms, to understand if there had been any issues which may have impacted the testing and to ensure that the context was taken into consideration. From this initial exploration, the preliminary frame for different aspects of the testing scenario was developed. During indexing, the draft framework was then applied back to all transcripts, and themes and sub-themes refined as ideas became clearer through further immersion. The researcher then charted the data to reduce material into understandable but brief summaries of what was said by participants (Ritchie et al. 2003). Finally, the data were synthesized by mapping and interpreting. Themes were checked against original transcripts and notes to ensure appropriate context and see if further changes were needed.

Data themes were organized based on the essence of the 5 research questions that were central to the aims of the cognitive interviews. Namely these were: students' response process; use of directions; preference for static vs simulation clusters; ease of navigation and use; and perceptions of TTS features.

The results are presented in accordance with the principle of 'thick description', to present give adequate voice to participants alongside the researcher's interpretation of these experiences. Quotes from participants are included in the report as indented, italicized quotations attributed only to a grade and school, to protect confidentiality. Quotes are provided verbatim wherever possible.

Results

1. To what extent do students need and engage with the directions provided?

Most students expressed a preference for directions, including all the students who used TTS. 6/11 Grade 5 testers and 9/11 Grade 8 testers preferred the simulation with directions (15/22 students overall, 68%). Of students who said they had no preference either way, most expressed in the interviews that while the directions could be helpful, they had already deduced the required actions independently, suggesting the need for a balance between guidance and autonomy. They also recognized that although they had been able to answer the question without directions, it could be helpful for other students.

- "It's giving me direction, to point me in the right direction of where I need to go. I think it was helpful knowing to let it go all the way through and not pause it at Generation 4, so I actually look at the graph." (9th grader, Timnath Middle/High)
- "It tells you specifically what to do, so you're not trying to figure it out. Like if there weren't directions, I would have definitely still figured it out, but I think it's easier with them to help you know what to do. Yeah, I think I understood them better when I read the directions for them." (7th grader, Lewis-Palmer Middle)

Students were also asked whether they would have answered the questions differently after seeing the directions. 13/18 students (72%) overall said they would not have given different answers, compared to 3/18 who said maybe and 2/18 that said yes, they would have changed their answers. However, even if they would ultimately answered the same, students noted that having the directions would have helped their cognitive process in engaging with the questions and problem-solving. For example, some students felt they could have reached the answer faster or would have considered the material more carefully with directions. It was noted in coding that after seeing the directions, some students were better able to articulate how they produced their answer.

- "I don't know if I would have answered them differently. I feel like I would have answered it the same, but this one just reminds me [to] make sure you really look at what the graph is telling you, what the table's telling you, so you can examine the table or the graph too." (7th grader, Timnath Middle/ High)
- "Maybe it would make me think more and try to check my answer." (4th grader, O'Dea Elementary)
- "It would have definitely helped me understand what I was doing in the test…and it definitely would help me understand what I was saying, but I don't think I would have changed my answers." (7th grader, Lewis-Palmer Middle)

Both step-by-step (numbered) and non-numbered directions were perceived as valuable comprehension aids. 20/28 students overall (71%) thought that the step-by-step directions were useful, including all students who used TTS. Several students noting that having these instructions was helpful in case they missed anything while working through the question; it served as a type of "check" over their work that aided comprehension and task completion.

- "I probably would have eventually, but I don't think I would have done it in that specific order...they helped not skip all the small things. With the shrubs, I didn't completely notice the shrubs until probably the second question. There's 3 graphs that I can read, so if I were to see this - "Examine the graph", I'd probably go back and examine all of the graphs." (7th grader, Skyview Middle)

15/19 students overall, (79%) said that the non-numbered information was helpful for them. Students who felt the information given was 'just right' thought it was helpful to contextualize what they were looking at.

- "It's just giving me an idea of what's going on, that way I can fully understand it before I answer anything…If it just had the numbered ones, I wouldn't understand. All that would tell me is to watch the animation, and I would have to try to figure out what's going on then by watching the animation instead of just reading that." (6th grader, Skyview Middle)

All students who said it was 'too much' information were testing the Grade 5 protocol. Some suggested that the question should only include details that the student definitely needs to know.

- "Maybe if you highlight big details that the student doing it should really know, and maybe you should put first, next, last so we know we're going to do this first, we're going to do that, and then we're going to do that." (4th grader, Mammoth Heights)
- "There's just a lot of extra sentences that you don't need in both the directions." (6th grader, PCK)

2. Do the assessment stimuli elicit the intended cognitive response process?

Grade 5 protocol – Simulation Question 1

Almost all students (15/16) showed engagement on the task, however only 3/16 got this question correct. 4/16 said that they had some familiarity with the topic already, and 8/12 students who were asked how easy or difficult it was to use the simulation to help them answer the question felt that it was 'easy' or 'very easy'.

Response times for this question ranged from 2 to 12 minutes. 2 students felt that they could not answer the question at all and did not record any response. However, most students could process the assessment stimuli and formulate their answer quite quickly and effectively (median response time was 4 mins), demonstrating effectiveness in understanding the task requirements. Students who more easily interpreted the task described the question in terms of observed change, and most referred to mass in how they understood what the question was asking.

- "It asks me to look at the water and see if it decreased from the stuff inside the thing, and then in Part 2 the balloon filled up because it was increasing." (4th grader, Mammoth Heights)
- "Asking about change in the mass in Part 1, and then to use Part 2 to explain what else is happening." (6th grader, Thornton Middle/ High)

However, for half of the Grade 5 testers, students did not appear to fully understand what was being asked of them. In 5 cases, the student was a 4th grader, and it appeared that the content may have been above their current cognitive abilities. All 3 students who answered the question correctly were 6th graders, 2 of whom were already familiar with the topic. When asked to articulate what the question was asking, these students' responses suggested confusion in identifying the key components of the question – i.e., that they needed to elicit information about *mass* specifically in relation to the *baking soda and vinegar* mixture and for the second drop-down, about change seen in Part 2 of the simulation. Many of the 4th grade students seemed to fixate on the image of the pennies, diverting their attention from grasping the essence of the question.

- "I think it's saying like, what does the penny look at first and then after?" (4th grader, Mammoth Heights)
- "Some vinegar, and salt and vinegar. How did they clean it if they're just that dirty?" (6th grader, Skyview Middle)

Students were asked to describe their cognitive process as best they could during the think-aloud. 14/16 were observed to be going back and forth between the simulation and the question to formulate their answer. Even when students did not seem able to articulate what the question was asking, they did appear to understand which of the simulation's components would help them answer.

- "It's [the data table] explaining what happened during the penny test, and how much the final mass of the mixture weighed." (6th grader, PCK)

The process students narrated in their think-alouds reflected how they used the appropriate parts of the simulation to formulate their answer. In some instances, the students described using only the data table to answer the question, and their answer appeared to be based on math reasoning and data interpretation alone rather than conveying a wider understanding of the scientific phenomena at play. Success on the item therefore provides less than conclusive evidence about student facility with one of the three cognitive processes intended for measurement.

- "I looked at the mass of adding baking soda and vinegar, and saw that 10 + 100 is 110, but the final mass was 105, so it decreases by 5. Interviewer: And what about the second part of that test question? "I looked at when they added baking soda and vinegar it was 110 and the final mass was 110, so it stayed the same." (6th grader, PCK)

Others described drawing on a combination of the animation and data table in the simulation, and their response indicated at least some understanding that the change in mass was due to the processes displayed in the simulation.

- "The baking soda might have been 10g in the beginning, but ended up making 105g because it dissolves inside of the vinegar - they do something, and the vinegar gets bubbly when you add baking flour because the baking flour had some type of chemical that makes it react like that, and the penny was also little cleaner…. [for the second drop-down] There was a bunch of bubbles because the chemical reaction of the baking soda entering the vinegar made a bunch of gas. The gas stayed inside, because of the balloon. So, I think that is evidence that mass just stays the same." (6th grader, Thornton Middle/ High)

Grade 5 protocol – Simulation Question 2

Students were quicker to form an answer to this question, likely because they were already familiar with the simulation, with response times ranging from less than one minute to 7 minutes, with a median response of 3 minutes. 7/16 students got this question correct, which was the highest number of correct responses seen in the Grade 5 testing. Of those who answered correctly, 4 were 6th graders and 3 were 4th graders. None had prior familiarity with the topic.

Just as in the first question on the simulation, it appeared that half of the students did not appear to fully understand what they were required to do to answer the question. Most students could articulate *what* the question was asking them; how to substantiate the claim about gravity made by the student in the simulation.

- "Out of all the statements, what statement best provides evidence to prove that the force of gravity can be observed during this investigation." (6th grader, Skyview Middle)

In most instances, their narration during the think-aloud revealed that they understood that looking for proof of gravity at work might mean looking for something sinking or being pulled down.

- "If it sinks to the bottom of the glass, then the gravity is pulling it down all right." (6th grader, Skyview Middle)

As in the previous question, most students also understood what aspects of the simulation would help them to answer the question – i.e. that the answer was contained in the animation rather than the data table. However, only 10/16 students replayed and interacted with the simulation before answering this question.

- “I used the picture as the picture from the simulation shows what's happening.” (6th grader, Thornton Middle)

The difficulty for students appeared to be that they struggled to understand *how* to use the simulation to help them answer the question, and their descriptions of how they formulated their answers also suggested that they were not necessarily able to connect what was shown in the simulation to the question asked.

- “I couldn't really pick an answer because the simulation didn't really represent any of those. I don't think A works, because I don't think gravity would affect that and then for B - I don't think that would work because it produces bubbles because of the chemicals. And then in C, I don't think that has anything to do with gravity. D, I don't think the soap sinks to the bottom of the glass because in the demonstration when they pour in the soap and then they stir it around, the soap bubbles are at the top of it instead of at the bottom of it.” (6th grader, Skyview Middle)

Three students felt that none of the multiple-choice answers connected to the animation. Where students did use the animation, their responses suggested that they struggled to identify a distinct instance of the force of gravity, and some felt that the animation showed all 3 substances equally sinking to the bottom of the glass.

- “They all sink to the bottom of the cup, but that's because they weigh very little because the baking soda is just powder. The salt doesn't really weigh that much, and soap just sinks down because it doesn't really weigh as much. I know that sometimes when stuff doesn't weigh that much, it might stay up floating. But sometimes that doesn't really happen, so I'm pretty sure that the baking soda and the vinegar mixture loses mass. Might be the answer, but I'm still kind of doubting myself.” (6th grader, Thornton Middle)

In instances where the student did not actively use the animation to formulate the answer, they instead relied on inappropriate reasoning to deduce an answer, for example, by referring to the previous question on pennies and mass.

- “I think it's the baking soda and vinegar mixture loses mass. I looked at the other questions and that was the one that makes the most sense. And I looked from the gravity pulling it down. I used when they're putting in the different - like dish soap, vinegar, into cups and taking the pennies out.” (6th grader, PCK)

Grade 5 protocol – Cluster Question 1

Response times on the first question of the cluster ranged between 1.5 to 9 minutes, with a median response time of 4 minutes. 9 students were asked to comment on how easy or difficult they found it to answer the question, of which 6 said it was ‘difficult’ or ‘very difficult’. Students were equally split on whether the topic was new or familiar to them, but it did not appear to be based on their grade level, as some students mentioned seeing similar experiments on YouTube or elsewhere online.

- "I'm pretty sure I already know what it's gonna happen because, I sometimes see when I use the Internet…I see some people adding oil to water and the water doesn't really mix with the oil. So, I know that adding some dye to it, like some watery type dye, it's not gonna mix up with the oil. Instead, it's gonna mix up with the water…" (6th grader, Thornton Middle)

In most instances (11/15) the student appeared to understand the essence of what the question was asking them, which they articulated as looking for a representation of the model setup in the passage. One student (who also was the fastest to answer the question) noted that what they were seeing in the model related to the concept of density, an above-level concept for grade 5, more complex than the answer the item called for. For some of the 4th grade testers, it seemed that the concept of a model was somewhat new to them, and two could not grasp the question at all. 4/15 students who attempted the question got it correct, equally split between 4th and 6th graders, only 1 of whom had prior familiarity with the topic.

- "I think it asks what materials would most likely, well most similarly, represent what's shown on the left." (6th grader, Skyview Middle)
- "Which model - like if you were to do this question, if you were to do this science project again, what else you could use in place of veg oil and water?" (6th grader, PCK)

Almost all (12/15) students went back and forth between the cluster and the question before answering, and understood that Part 1 of the passage was the relevant area to focus on. In their think-alouds, most students described mainly using the images in the passage to help formulate their answer.

- "Water, vegetable oil, clear marbles, yellow marbles… I'm thinking B would currently have the most difference, but also clear marbles and pepper, pretty big difference. So, B would be the most like the diagram, at least I think so that's why I chose B."(4th grader, Mammoth Heights)

However, in some cases the images confused students – particularly the "gentle swirling" of the bottle in the model, which some students interpreted as bubbles, and incorrectly inferred that the mixture was causing a chemical reaction to take place.

- "I think it's A because it has clear marbles that looks like the food coloring, and the yellow marbles are like the balls in the in the bottle that are getting sent to the top." (6th grader, Skyview Middle)
- "I'm not really sure which one would be the best answer for deciding between B and C. I think if you're trying to represent the bubbles in the bottle, you would want marbles to represent that, and D is just pepper and salt so I don't really think that would be a good answer." (6th grader, Skyview Middle)

Although most understood that they were looking for a representation, sa number of students seemed unsure about which properties of the mixture held greater significance for accurately representing the content of the model. For example, some described looking for a direct likeness. 4/13 students who answered the question focused on the color of the different components, noting that both vegetable oil and one set of marbles were both yellow.

- “I know D isn't the answer because I know that answer isn't the answer, because water and pepper, they didn't add pepper, vegetable oil and salt. They didn't add salt so that only leaves us with one more answer, which is answer A - vegetable oil, yellow marbles and water, clear marbles.” (6th grader, Thornton Middle)
- “Um, pepper and clear marbles. Because the pepper is dark, and it could be like the food coloring and the clear could be like the stuff around it.” (4th grader, O’Dea Elementary)

Grade 5 protocol – Cluster Question 2

12 of 16 students attempted this question, with some needing to skip due to time constraints of the interview. No student answered this question correctly. 10 students were able to articulate in their own words that the question related to the weight of the materials before and after the investigation.

- “It's asking what are the weights of the full thing, and then what are the weights of it after this [the tablet] dissolves.” (6th grader, PCK)
- “It's explaining that in Part 2, the students decided to weigh all the materials before they dropped the tablet into the bottle and sealed it, and so now I think it's trying to tell me to combine the weight of all the materials to see what the ending mass would be.” (6th grader. Skyview Middle)

All students understood that they needed to use the information in the table in the question to formulate their answer, and that they were looking to add or subtract based on total combined weights before and after the investigation.

- “I think I need to measure the total; I mean I need to weigh the tablet and cap. I find the information right above it, it says 'Weight before investigation'.” (6th grader, Skyview Middle)

5/12 students moved back and forth between the images in the passage and the question to come up with their answer, but 4 noted that they had not used anything from the passage to answer the question, relying on the data table alone.

- "I didn't really know what was going on until I looked at the key. And on the right side, it was kind of confusing what they were saying, but once I saw the little table, it made a little bit more sense what I was meant to do." (4th grader, Mammoth Heights)
- “I didn't like how in Part 2, the information isn't needed as much as Part 1, and you don't really need the Part 2 pictures." (6th grader, PCK)

For several students, the diagram in the passage caused greater confusion, as they felt that the actions represented in the passage should have impacted the weights but did not know whether this information was relevant and/or how it should be applied to answer the question.

- “It asked me how the weight changes when they add the tablet, and how the cap and the tablet have things in common, because the bubbles can have gas in it, and the bubbles can have weight.” (6th grader, Thornton Middle)

- "You have to think about this one. For the tablet is that - is the 10 grams, is it the full tablet or is it 1/2 of the tablet because they broke it into 2 pieces?" (6th grader, PCK)

In all, the aspect of this question that caused greatest difficulty for students was understanding what they needed to do to answer the question. Students' narrations in the think-aloud showed that they understood they needed to drag the bars up, but they did not understand how to do so given the graph shown. 8 of the 12 students who attempted the question needed prompting or guidance from the interviewer on what they needed to do.

- "I'm looking over all of the animations, I don't think the directions really make sense." (6th grader, Skyview Middle)

Grade 8 protocol – Simulation Question 1

Almost all students (9/11) showed engagement on the task, however one student noted that they could not answer the question at all and did not record any response. For those students who did attempt to answer, 4/11 got the correct answer, all of whom were 7th graders. Response times for this question ranged from 5 to 10 minutes. The median response time across students who did provide an answer was 6 mins. 5 students said that they had some familiarity with the topic already, of which 2 answered the question correctly. 7 students said it was 'easy' or 'very easy' to work with the simulation to answer the question.

Students generally understood the concept of advantages or disadvantages based on fur color in survival situations, and the relevance of adaptation and reproduction in determining survival rates. In their narration, several students also indicated that they understood the importance of the environment or site characteristics to the question.

- "I thought this question was asking me to figure out in these climate zones, which one would have a better advantage depending on the color of their fur and how they live." (7th grader, Skyview Middle)
- "I think it's asking what would these - how does the rate of survival and reproduction at this one site change over time with certain conditions in the place that they're in? And if it's a disadvantage or an advantage for the different type of fur color." (7th grader, Timnath Middle/ High)

9/11 students went back and forth between the simulation and the question before answering. While some students seemed unsure initially or needed clarification on the task, others demonstrated a clear understanding of what was being asked and which information to consider. However, 7 of the students showed initial confusion in understanding what was expected of them to answer the question, with the main difficulty being that most did not realize at first that they needed to use Part 2 to answer the question. These students required prompting from the interviewer to find and engage with Part 2 of the simulation.

- "I clicked the buttons, the little circles that you told me about, and then I clicked play. That helped me out. I didn't realize that was there until you pointed it out." (7th grader, Skyview Middle)

In their narration, several students noted how the simulation helped them visualize what the question was asking. However, others struggled to connect what they saw in the animation and data table to what was being asked. The dropdown options seemed to help some students understand the context better. One student noted that the misalignment between the fur colors referred to in the simulation (dark brown, brown, gray) and the question text (light brown, light gray, dark gray) was confusing for them. The stimulus for the item intentionally incorporated these new colors to support inference (i.e. the seeming misalignment was deliberate rather than an oversight).

- "What helped me was just whenever I played this simulation, I would notice how over time the jackrabbits with light gray fur would start decreasing over time, and I assume that that would be because of the drought that it was talking about." (7th grader, Timnath Middle/ High)
- "I had a general understanding of how you were supposed to do it, and then the physical characteristics helped me a lot to give my answer." (9th grader, Timnath Middle/ High)

Regardless of whether they answered the question correctly, it seemed that students were using a mix of observation of the simulation and critical reasoning to answer the question. Most students were able to vocalize how they compared the advantages and disadvantages of different fur colors in Site 1, considering factors such as survivability, reproduction rates, and adaptation to environmental conditions. Six students' narrations showed that they considered and combined what they knew about various environmental or scientific factors such as temperature, camouflage ability, heat absorption, and habitat suitability in their reasoning.

- "Jackrabbits with light gray fur would have a disadvantage, because when drought starts to happen, all the color of trees and branches start to turn brown as they start to die. So, I think the rabbits with gray fur won't be – There won't be as much shade that the plants would produce, because there's not many plants. And then Site 1 - compared to jackrabbits with gray and dark fur, jackrabbits with brown fur would have an advantage because all the soil and everything is turning brown and getting really dead, if that makes sense." (7th grader, Lewis-Palmer Middle)
- "… I saw the physical characterizations, and then the fur color and their fur length and I was thinking if one site's really hot, they probably would have short hair to live, because otherwise it would be way too hot for them to survive, right? But then for Site 1, it said it's gonna be a variety of temperatures and stuff. And I would say long would be the best for that because it could either be really hot or really cold. But like, it's probably not gonna be as hot as Site 2. So, I feel like the long-hair would be a lot happier and the advantage at Site 1, because if it's like really cold, the short-haired, the short fur length in the gray would not have an advantage. They would have a disadvantage at Site 1." (9th grader, Timnath Middle/ High)
- "Darker colors seem to absorb more heat. So, I think if you had a lighter color of fur, then it would reflect more of the heat so you wouldn't get as hot. The darker gray fur, it's darker than the lighter fur obviously, so I think that it would absorb more of the heat, and it would end up causing the rabbits to overheat at one point. Especially in a drought, when there's less water around." (7th grader, Skyview Middle)

Grade 8 protocol – Simulation Question 2

As in the Grade 5 protocol, students showed quicker response times on the second question of the simulation. Out of all the test questions, this question saw the highest number of correct responses, with 9/12 students providing the right answer. Of the students who answered incorrectly, two were using TTS. Response rates for this question ranged from 1.5 to 7 minutes, with a median response time of 3 minutes.

All the students attempted to answer the question, and generally seemed to understand what was being asked of them. All of them were able to articulate that the question was asking them about change in the population from Generation 1 to Generation 4, but some students struggled to interpret what the second drop-down in the item was asking of them.

- "This question is asking me in my own words how much did - Over generations, how much did the population grow of gray fur jackrabbits, and then also why do you think they grew over time?" (7th grader, Lewis-Palmer Middle)
- "I think it's asking that in the second part of the simulation, how many more of the gray fur rabbits are there than the brown fur." (7th grader, Lewis-Palmer Middle)

Most (9/12) students were observed to move back and forth between the simulation and the question before finalizing their answers. Students also easily recognized the need to refer to the simulation to calculate percentages and understand the population dynamics between different generations. Several students recognized the relationship between the visual representation in the animation and the numerical data provided in the data table, using both to support their understanding.

- "So, I can see kind of visually and get a broad picture. And then at the graph [data table] below I can see a more in-depth - 'Hey this is what you just saw' – It's what the graph kind of shows from the simulation." (7th grader, Lewis-Palmer Middle)

Others felt that they did not really need the animation in the simulation to answer the question but could base their answer on the data table alone.

- "The simulation is actually not that helpful, because we have the table under here and we can just use the data directly." (9th grader, Timnath Middle/ High)

In terms of formulating their answers, students varied in their approaches, using a mix of observation, deductive reasoning, and inference to respond to the question. About half noted that they drew on what they observed in the animation and the data in the table to come up with their answer.

- "I went over here [to the data table] and then I used background information from gray fur. And I looked over here and I looked at the gray fur and then I saw that it changed. I went over here [to Generation 4 in the animation] and it looks like there's way more gray fur than brown fur." (9th grader, Timnath Middle/ High)

It is worth noting that 7/ 12 students did not calculate the percentage increase based on the table, but instead observed that there was a visual increase in gray fur jackrabbits, and deduced that the correct answer must be 80%, since that was the only answer option greater than the 50% originally observed in the population.

- "I'm just not good at knowing percentages right now, so I don't know how to find out that percentage, but it already started as 50% and then the gray fur amount has gone up. So, it just kind of seems that would be 80, but I'm not completely sure…I was guessing since like there's a larger population of gray fur." (7th grader, Timnath Middle/ High)
- "They had more because they reproduced or they put in more, and so I thought that it would be 80% because if it was more than 50, if it had to be." (7th grader, Skyview Middle)

Most students noted in their narrations that they interpreted this change between generations as an indication of gray fur providing an advantage to the population at Site 2. For this item, fewer students articulated any reasons for why this change might have occurred. 3 students expressed uncertainty about their answers, acknowledging that they were not entirely sure of the reasons behind their choices. Those who did voice their reasoning were able to connect the question to what they understood about adaptation and survival.

- "I was guessing since there's a larger population of gray fur. I'm not completely positive why I think that there's an advantage, but I did notice that brown fur went down when gray fur went up, but at the same time brown fur went down. So, I'm not sure if that's an advantage or disadvantage, but I would assume it might have been an advantage." (7th grader, Timnath Middle/ High)
- "I'm gonna say provided an advantage for this one, because it looks like they have like a better chance of living in the color environment that they're in, cause like the color of their fur goes better with this color." (7th grader, Lewis-Palmer Middle)

Grade 8 – Cluster Question 1

4/ 11 students answered this question correctly, of whom two were using TTS for the question. This percentage is similar to the p-value when the item was Field Tested (.42). All but one student appeared engaged while working on this task. One student felt that they could not provide an answer to this question. 5 students noted that it was a new topic for them, and of the 9 students who responded, 5 felt it was 'very difficult' or 'difficult' compared to 4 who said it was 'easy' or 'very easy'. Response times on the first question of the cluster ranged from just under 2 minutes to 7 minutes, with a median response time of 3.5 minutes.

6 of the 11 students who worked on the question appeared to struggle to understand and articulate what was being asked of them. Most could understand that the question focused on their interpretation of the reproductive process depicted in Figure 2, including its mechanisms, requirements, and outcomes, even though few described what was being asked in terms from the passage like cross-pollination or fertilization.

- “I think this test question is just asking me about a second reproduction process and it's just asking me about what are some of the like, how does it happen? Does it happen with 2 plants or one plant, and then also what does the reproduction do.” (7th grader, Lewis-Palmer Middle)

Students seemed less certain in articulating what the second and third items were asking around the concept of genetic material and its implications in reproduction. 3 did not understand what the word “combining” meant in this context i.e. the combining of genetic material.

- “I think it's asking me about how this plant grew and what it needed. For this one, I don't know how to word this last part - with the identical part - but I think it's asking if it most likely looks like the seed that came from it.” (7th grader, Lewis-Palmer Middle)
- “I know what all the words mean, but I don't understand what 'No combining of genetic material' means, cause I feel like if there was no combining of genetic material then the plant just wouldn't be able to grow. I'm probably wrong, but I feel like that's how that works.” (7th grader, Skyview Middle)

To formulate their answer, most students were able to infer that since the process starts with a single seed and involves the growth of a single plant, it only requires one plant. Some also referenced Figure 1 to distinguish this process from those that require the combining of genetic material from multiple plants.

- “I looked at Figure 1, and I realized it has to have 2 plants to make another plant. But in Figure 2, it needs one plant because it doesn't use pollination. So, it doesn't need 2 plants to make it. And then because the first question of 1 plant, it doesn't need combining like they did in Figure 1, of combining 2 plants to make another plant.” (9th grader, Timnath Middle/ High)
- “I think it's one plant because the picture is showing the same plant growing and growing throughout the time. I think it's not combining because in Picture 1, it shows one plant and 2 plants trying to make, I guess this thing. I don't know what that is, but there's 2 of the same thing combining to make something else. But in the second one, there's only one plant growing and growing.” (7th grader, Skyview Middle)

All students appeared to understand the task requirements, noting the importance of Figure 2 (versus other sections) as relevant to answering the question. Most understood that they needed to analyze the visual information presented in Figure 2 to make inferences about the reproductive process, even if they had difficulty discerning the exact relevance of Figure 2 to answering the question.

- “It gave us the information and it gave us a very clear and straightforward picture of how it’s growing, and then it just keeps going through the cycle. And then when it tells you to look at Figure 2, that gives you more information and it tells you like, ‘Hey, the answer’s right here’, it’s not in the whole paragraph.” (7th grader, Skyview Middle)

Grade 8 – Cluster Question 2

9/ 12 students attempted the second question of the cluster. Students appeared to show less confusion or difficulty on this question than the previous, and all those who attempted this question showed engagement on the task at hand. However, only one student answered this question correctly. In field testing 40% were able to answer this question correctly.

All students could articulate in their own words that the question was asking about the change seen in the dandelion population after a period of using a chemical in a field.

- "It was asking so if you used a chemical to kill dandelions in a field and then after 4 months you decide to use it again, how would it affect the dandelions?" (9th grader, Timnath Middle/ High)

Based on the information in the passage, most students were able to infer that the chemical had a negative effect on the dandelion population, but for 2 students this was less clear, and they felt that the outcome was more ambiguous.

- "[It's asking] if the dandelions grew better with or without the chemical, and if the chemical helped them grow or not." (7th grader, Lewis-Palmer Middle)

In considering the impact of the mutation on the dandelions, none of the students attributed the cause of the mutation to the chemical used. Some speculated that the dandelions could even grow resistant to the chemical over time.

- "Even after the dandelions were subjected to the chemicals they grew used to it, and since 50 dandelions were only allowed there at the beginning anyways, I think it's safe to say that after these dandelions grew mutation to not be so affected by the chemical, I think they could just go back to normal, even with the chemicals." (9th grader, Timnath Middle/ High)
- "I saw it [the data table] saying how much dandelions were there, but realizing if it's like growing back that fast, I feel like that's how I got my answer. It probably is mutating to be resistant to the chemical then, just not like before." (9th grader, Timnath Middle/ High)

One aspect of the question that caused apparent difficulty for some students was ambiguity around the word '*beneficial'*. 3 students reasoned that benefits incurred could relate to the dandelion population or to the humans who might benefit from the field being clear of weeds.

- "For this mutation, it's asking is it beneficial for - It's beneficial for the humans because they don't have any weeds on their field anymore, but it's harmful to the plant species if that makes sense." (7th grader, Lewis-Palmer Middle)

Almost all students (7/9) were observed to go back and forth between the passage and the question text before providing their answer, and 8/9 could easily find the information needed to answer the question. Some narrated that they were using both the images and the data table to come up with their answers, whereas for others it seemed that they were using the images only.

- "So, when I'm looking at the figures, I see that this is before [the] chemical product is used, but then after it is most of the dandelions are completely wiped out, but some of them begin to grow. Yeah, here I can look at the table as well to see." (9th grader, Timnath Middle/ High)

3. Do students find simulative clusters more, less, or equally engaging to static clusters?

Students showed varied preferences for simulations versus clusters, with factors such as comprehension clarity and visual representation influencing their choices. At all grade levels tested, there was no definitive trend in whether it was easier to find the information needed with the simulation vs. the cluster. The findings were similar for the students who used TTS; 3 found the cluster easier, 1 found the simulation easier, and 2 said they were the same. No clear preference was seen based on students' score level in previous CMAS testing either.

Analysis of students' views on the simulation or cluster showed that their reasons for liking one or the other were similar regardless of their stated individual preference. In sum, students liked when information was presented in a simple, concise manner without them needing to worry about misconstruing information or making inference errors.

Those who felt it was easier to find information in the simulation liked that it was interactive and visually engaging, and that it shows complex phenomena in a more dynamic way compared to a static cluster. They felt this made it easier for them to grasp the key information needed to answer the question.

- "It was very simple to understand what I needed to do with it, and I just pressed a few buttons." (9th grader, Timnath Middle/ High)
- "I think it was easier to use the simulation than the passage, because the simulation showed what was happening. Sometimes I feel like I thought something else, different words than what I meant to, in the passage, and then I got the wrong idea." (7th grader, Skyview Middle)
- "The pictures were harder because with the simulation you're actually seeing what's happening. But with the pictures, you kind of have to assume some stuff is happening." (4th grader, Mammoth Heights)

Those who found the cluster easier felt that the information was organized and presented in a more straightforward way compared to the simulation, which made the question easier to answer.

- "I think the passage because there wasn't so many pictures and stuff, and it was kind of straightforward. The other one, the bunny one had a lot of pictures everywhere, and I guess the pictures were helpful, but there was a lot of them and a lot of - I don't know, I guess the dandelion one was more straightforward, and it made it easier for me." (7th grader, Lewis-Palmer Middle)

- "I think it's easier for a reading passage because the simulation doesn't have words in it. Sometimes kids think totally different things than what it's like showing you." (4th grader, O'Dea Elementary)
- "It wasn't as confusing to me. It broke down the question more, and so it was easier to understand an answer." (6th grader, Skyview Middle)

Students' stated preferences were then compared to their observed thoughts and behavior during the think-alouds. Differences were seen between the grade protocols based on how students showed engagement, confusion, or difficulty during the think-alouds when interacting with each cluster type. Regardless of their stated individual preference:

- More students using the Grade 5 protocol showed engagement and less confusion/ difficulty working with the simulation than the cluster.
- Conversely, for the Grade 8 protocol, more students showed engagement and less confusion/ difficulty using the cluster compared to the simulation.

This discrepancy between students' stated preferences and observed behaviors during testing suggests a more nuanced pattern of engagement and interaction, and the importance of age-appropriate assessment stimuli.

4. What is the level of difficulty or ease in navigating the different types of stimulus information to answer the associated questions?

While some students seamlessly engaged with the stimulus information, others required guidance to navigate CMAS materials effectively. In general, students across all grade levels showed similar patterns of navigation; they had most difficulty navigating Sim 1 Question 1, but their need for guidance declined as they worked through the test questions, to needing virtually no assistance to navigate the final question using the cluster. The following paragraphs present the results for each grade level and cluster type separately to aid understanding.

Grade 5 protocol - Simulation cluster

For both questions, most students read the question first rather than navigating straight to the simulation. Across both questions in the simulation cluster, there were 5 instances of difficulty observed that the coder was at least somewhat confident related to navigation issues, where students got stuck because they struggled with how buttons worked – 4 of these occurred in Q1 and one for Q2. These students required a steer from the interviewer to find Part 2 of the question, or to engage the Play button. Further analysis suggests that in some instances it seemed that students were hesitant due to the testing setting and felt that they needed permission to proceed from the interviewer, or that they had simply not noticed the relevant buttons at first, rather that it being the case that they didn't know what these buttons did.

Including those who were prompted, all students used the Play button and about half replayed the simulation at least once before answering the question. Students' prior experiences with other online platforms influenced their navigation strategies. When asked, students said they intuitively understood what the play and replay buttons meant from other digital experiences.

- "Usually when you're like watching a series or something, there's always that button to start it." (6th grader, PCK)
- "If you know me, I use my phone a lot, so I know what a play button looks like because I use a bunch of TikToks." (6th grader, Thornton Middle)

Similarly, students appeared to intuitively know how to use the drop-downs in the items. One student noted being familiar with the format from similar tests she'd taken before. 2 students also used the tools above the simulation (calculator, ruler, etc.) to annotate the question, with one explaining that her math teacher had shown the class how to use these features before last year's CMAS.

Grade 5 protocol – Static cluster (Cluster)

Students showed different navigation patterns between question 1 and 2 of the cluster. For the first question, most students (13/16, 81%) started with the cluster first before moving to the question. On the second question, more students started by reading the question first (9/16, 56%, looked at the question first). However, this behavior is not unexpected given that the material in the passage had not changed between the 2 questions.

Grade 5 testers showed more confusion and difficulties with navigation using the cluster compared to the simulation, with all but one instance relating to the second question, which required them to drag bars up a graph. 7 students had some difficulty understanding how to drag the graph bars up to answer the question. Some of the 4th grade testers also had trouble controlling the mouse to do so. One student noted that it was "so annoying" to use, and another suggested that what students were expected to do with the graph should be clearer.

- "Maybe you could put a little note by the graph or something and say, 'Drag these lines up to do that' or whatever." (4th grader, Mammoth Heights)

Grade 8 protocol – Simulation cluster

Unlike the Grade 5 testers, students interacting with the Grade 8 simulations showed different navigation patterns between each question. For the first question, 7/12 students (58%) interacted with the simulation first when looking at the task. For the second question, almost all students (11/12, 92%) navigated to the question text first when interacting with this item.

There were 7 instances of confusion or difficulty that the coder felt confident were related to navigation issues. All but one related to the first question. Additionally, all 3 students using the TTS needed guidance from the interviewer on how to access the actual question to answer once the TTS had finished speaking. On the first question using the simulation, half of the Grade 8 testers struggled with how the buttons worked, including all 3 students who were using TTS. These students needed guidance from the interviewer to complete the question. The difficulty

seemed to stem from 2 places in general. Firstly, some students failed to recognize that they needed to navigate to Part 2 of the simulation to answer the question. Secondly, some did not understand from the simulation that they needed to toggle Site 1 or Site 2 in Part 2 to activate the Play button.

- "Oh, I thought those were the different questions, I didn't know they went with this. So does Part 3 go with it too?" (7th grader, Lewis-Palmer Middle)
- "I was actually confused because it doesn't tell me what to do. I just clicked it and kind of figured it out, but it would be much easier and much nicer if it tells you to click that button, it will unlock you more data." (9th grader, Timnath Middle/High)

By the second question in the simulation, students seemed to have figured out how to navigate the simulation with far less difficulty reported. They were easily able to describe how they could use the simulation to engage with the second question.

- "All of the buttons are right there, and simple, not confusing." (7th grader, Lewis-Palmer Middle)
- "You can switch from Site 1 to Site 2, and it has a table with the numbers you need to answer the question." (7th grader, Timnath Middle/High)

Only one student showed difficulties related to navigation for this question. This student very quickly paused the animation as soon as it reached Generation 4. This meant that the animation did not play through for the data table to populate, and so she had difficulty answering the question.

Grade 8 protocol – Static cluster

In contrast to the simulation, navigation of the cluster seemed to be a lot more straightforward for students testing the Grade 8 protocol. For the first question using the cluster, 9/11 (82%) started by navigating through the cluster first, but on the second question their behavior was more mixed, 6/11 (55%) started by reading the cluster first, compared to 5/11 (45%) who started with the question. Analysis of the recordings did not suggest any navigation issues relating to the cluster, with all students looking at Parts 1 and 2 of the passage unprompted. Students described the cluster as being quite clear in what they needed to do, including for TTS users.

- "It's very deliberate about what I had to do. I had my question on the right and I had 2 different figures about how plants could reproduce, and at the top it there were instructions of what I should do, and I also had pictures. So, as I was reading or as I had my Text-to-Speech read that to me, I could see the picture and see what it was talking about and see what it was explaining about." (7th grader, Lewis-Palmer Middle)

5. Do the accessibility tools (Text-To-Speech) help or hamper student's ability to answer the questions?

Almost all the students using TTS were already familiar with all the on-screen controls, were comfortable using the feature's functions, and were able to navigate the assessment stimuli as needed with minimal guidance from the interviewers. One student was not sure if they had used TTS before, and another noted that although it is available to him, he doesn't always use it. No student using TTS got more than one question correct, however this is not dissimilar to students who took part in the testing and did not use TTS.

Students varied in their use of TTS, with some choosing to use it to read through the stimuli and questions, while others did not immediately employ it. For 3 of the 4 questions presented in testing, most students started by using the TTS for the stimulus, and then moved to the question. Some students stopped or paused the TTS at certain points, either to reflect on what was being read or to navigate to different sections of the stimulus.

Based on their on-screen behavior, it appeared that some students seemed to struggle to auditorily process all the information they were hearing from the TTS at once, although, none of them attempted to slow down or change the TTS speed at any point. In these instances, and others, TTS was used iteratively, with students replaying specific sections to clarify understanding or find relevant information. Several students who used TTS appeared to have difficulty understanding where to direct their attention when the TTS moved between different parts of the stimuli presented.

- "The figures were a little bit more confusing for it. I think it went to Part 2 on the other side when I was there, I don't know where it went." (7th grader, Timnath Middle/ High)
- "It was hopping around a lot and there was a lot of talking at once." (7th grader, Lewis-Palmer Middle)

Half of the students using TTS were also not sure where or how to find the items they needed to answer, as the picture descriptions used by the TTS were on the right-hand side, where the question would normally be.

- "For me it was kind of just like where to click sometimes. Well, I guess cause the first run-through, I thought there was going to be questions, but then I realized it was all just reading." (7th grader, Lewis-Palmer Middle)

Students had mixed views when asked to share their perceptions of the TTS and whether there was anything confusing about the feature.

- "I think it was very good how it was set up, and I think it was really easy to use it to access. It was really nice, and I liked how it was separate from the top bar, I really liked that." (7th grader, Lewis-Palmer Middle)
- "The text to speech could be a little bit better. It got a lot easier, but at the beginning it was super hard because I didn't know how to stop it, or if it kept on going and I didn't know where it was going." (7th grader, Timnath Middle/ High)

Conclusions

1. To what extent do students need and engage with the directions provided for simulations on the CMAS science assessments?

While all students were able to complete the simulations without directions, most students did suggest that directions would be helpful, particularly those using TTS. Pearson recommends that directions are kept but streamlined. Pearson also recommends moving the simulation directions above the simulation, and anything else previously contained on the direction asset (on the right side of the screen associated with the simulation) to a tab (on the left side of the screen as a stimulus where the student could access it at the same time as an item).

2. Do the assessment stimuli elicit the intended cognitive response process?

Most students at all grade levels quickly processed the assessment stimuli and answered within 3-4 minutes, indicating a good grasp of task requirements. However, about half of the students, especially 4th graders, struggled to understand how to answer the questions. With the Grade 5 protocol, some students showed difficulty connecting what they were seeing in the simulations to the items. For the Grade 8 protocol, students generally understood how to combine the simulations and data to answer the items, but in some instances were relying on the drop-down options to formulate their answer rather than applying analytical skills to the clusters. Improving the clarity of directions given should help to enhance their understanding and support students to make connections between the stimuli and scientific concepts.

3. Do students find simulative clusters more, less, or equally engaging to static clusters?

Students' opinions were split on this topic with some preferring the animations and others preferring static images. Pearson recommends continuing with both modes of presentation.

4. What is the level of difficulty or ease in navigating the different types of stimulus information to answer the associated questions?

Students' opinions were also split on this topic. However, their navigation behavior in the cognitive labs showed they more easily navigated the static clusters. This may be partially due to the lack of directions given prior to the stimulus, and because once they had progressed through one cluster, they were more familiar with navigation (the static cluster was always presented second).

One thing that seemed to be difficult for students in both presentations and at all grade levels was understanding the naming conventions used on the tabs on the left side of the screen that contain the stimuli. CMAS Science references these tabs as "Part 1", "Part 2" etc. Students often did not associate these "Part" names on the tabbed stimuli as providing direction to the student as to where to navigate to help them find answers. When directly asked about what they thought reference to "Part 1" and "Part 2" meant, several students responded that they thought it meant to do things in a specific order like "Step 1" and "Step 2". The students in the cognitive labs referred to the parts by their contents like "the video with the pop bottle" or the "investigation with the baking soda". Pearson recommends mimicking this behavior and changing the tab

names to more specifically reference the content of the tab. Similarly, references to the simulation/animation should call it a “video”.

5. Do the accessibility tools (Text-To-Speech, TTS) and features (design of art and layout) produce an inadvertent effect on the construct assessed? i.e., Does the tool help or hamper student’s ability to answer the question?

The students found the TTS directions asset and TTS for simulations confusing. Historically CMAS SC has used an item level asset (at the right side of the screen) with screen captures from the simulation to indicate what the words were in the simulation. Most students did not understand this presentation, and many skipped it entirely. Students were also unlikely to follow along with the on-screen reading of the TTS, as it reads all tabs associated with an item - even if those tabs are not needed to answer the specific question the student is answering. A student must click on each tab separately in order to follow the TTS as it reads.

Pearson recommends moving as much of the TTS as possible into the stimulus asset. To achieve this, audio recordings of the text in the animations will be synced to the animation and the animation highlighted to show the progression of the audio through the simulation. A technology enhancement request has also been made to prevent TTS from reading through all the tabs automatically, and only reading the active tab (the tab the item is associated with).

The item level and static cluster TTS was easily navigated by students, and no changes are recommended to those assets.