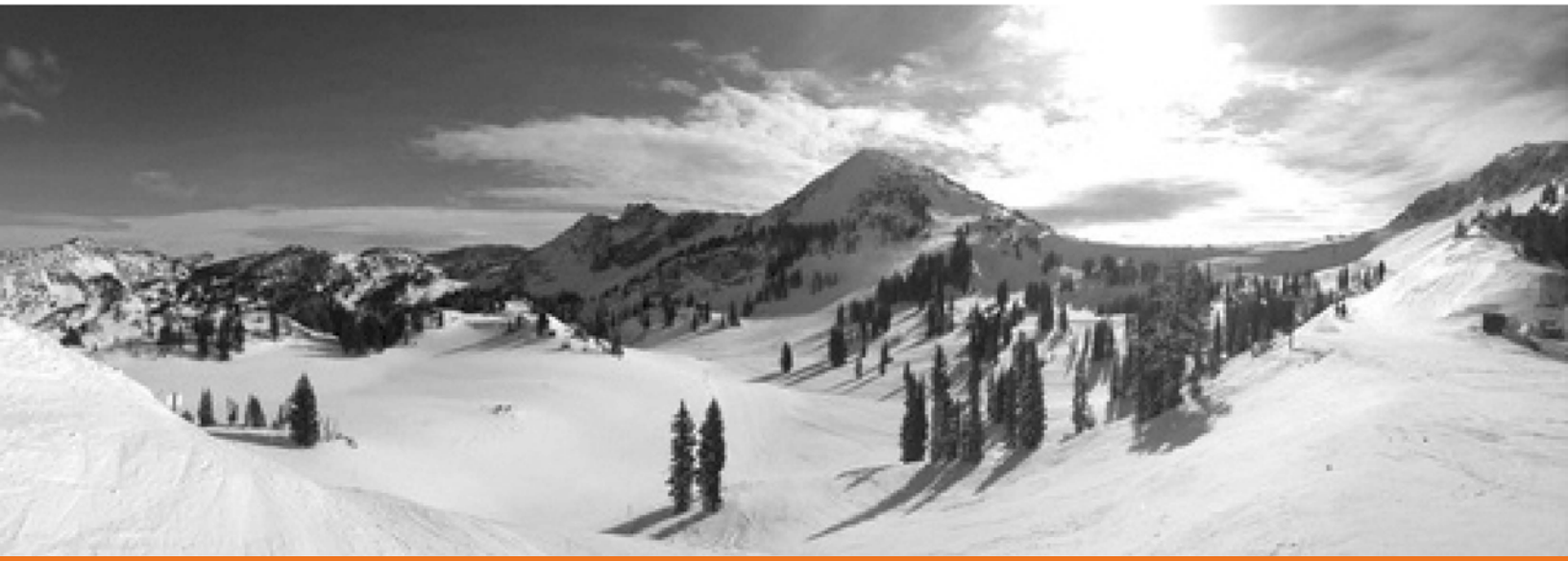
Colorado Measures of Academic Success



Technical Report

Math, English Language Arts (ELA), and Science

2022

Colorado Measures of Academic Success (CMAS)
Technical Report

2021–2022

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, content, administration, and technical characteristics of the Spring 2022 Colorado Measures of Academic Success (CMAS) assessments for mathematics and English language arts (ELA) in Grades 3–8; for science in Grades 5, 8, and 11; and for the Colorado Spanish Language Arts (CSLA) assessment in Grades 3 and 4. The report includes an overview and summary of the components of the program, including information regarding the planning and administration of the assessments and details regarding item development, test construction, administration procedures, scoring, reporting, reliability, and validity, as well as a statistical summary of the Spring 2022 operational and field test items.

1.1. Testing Requirements

All public schools in Colorado are required by state law to administer a standards-based summative assessment each year in specified content areas and grade levels. Every student, regardless of ability or language background, must be provided with the opportunity to demonstrate their content knowledge through the state assessments. The CMAS assessments in mathematics, ELA, science, and social studies are Colorado's end-of-year standards-based assessments designed to measure students' achievement of the grade-level or grade span Colorado Academic Standards (CAS).

As a requirement of Colorado School Law C.R.S. §22-7-1006.3 (4) (a) and (b), English learners with Spanish as their home language in Grades 3 and 4 who meet established eligibility criteria may take the CSLA forms of the CMAS ELA assessment. The CSLA forms serve as accommodated versions of the CMAS ELA assessments and are parallel and comparable to CMAS ELA in test design, item type, scoring, and reporting.

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 selected by local educational providers to be administered to their students. The comparable paper-based forms may also be administered to students with disabilities and English learners as appropriate in schools that otherwise are administering the online forms of the assessments.

In 2015, Colorado passed legislation (C.R.S. §22-7-1013 (8) (a-c)) that allows for parents/guardians to excuse their child(ren) from testing.

1.2. Intended Population

The CMAS assessments are intended to be taken by all students enrolled in public schools, with the exception of some students with the most significant cognitive disabilities who take the Colorado Alternate Assessment (CoAlt) assessments as determined by the student's Individualized Education Plan (IEP) or other educational team. English learners in their first year in the United States are exempt from the ELA assessment. However, English learners in Grades 3 and 4 designated as not English proficient (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. Students with disabilities and English learners 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.

1.3. CMAS Background

The CMAS Science assessments were first administered in 2013–2014, the CMAS Mathematics and ELA assessments were first administered in 2014–2015, and the CSLA assessments were first administered in 2015-2016. Colorado developed the CMAS Mathematics and ELA 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.

In 2017, the State Board of Education provided direction to the Colorado Department of Education (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 (including the CSLA) previously administered in Colorado to maintain longitudinal trend data. Test forms based on the abbreviated blueprints were developed in Fall 2017 and administered beginning in Spring 2018.

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. Students were required to take one test in either mathematics or ELA, depending on their grade, although parents/guardians could choose to have their children take both tests. With the exception of students with a parent/guardian excusal, students in Grades 4, 6, and 8 were required to take the mathematics assessments; students in Grades 3, 5, and 7 were required to take the ELA assessments; and students in Grade 8 also took the science assessment. The Grade 5 and high school science and Grades 4 and 7 social studies assessments were not administered.

In 2022, newly revised standards were implemented for mathematics, ELA, and science. In 2008, Colorado passed Senate Bill 212 (also known as CAP4K) 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. While minimal changes were made to the mathematics and ELA CAS, the science CAS underwent a substantial update to keep up with the shift to the Next Generation Science Standards (NGSS; NGSS Lead States, 2013)[2]. Full implementation of the new science standards took place in 2021–2022.

Regular testing procedures resumed in Spring 2022 for mathematics and ELA (including CSLA). Colorado students saw items aligned to the 2020 Colorado Academic Standards (CAS) for the first time in spring 2022. The new science assessments based on the new three-dimensional science standards were administered to all tested students which a made it possible to test enough

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

[2] Next Generation Science Standards is a registered trademark of WestEd. Neither WestEd nor the lead states and partners that developed the NGSS were involved in the production of this product and do not endorse it.

new content to allow for a robust item bank and to obtain a sufficient number of students to conduct field test analyses. Social studies was not administered in Spring 2022.

1.4. Purpose of CMAS

CMAS assessments were designed to be used for a variety of purposes, including informing parents 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. Results should be taken into consideration alongside other achievement information available locally. Results are also used as a piece of information in the evaluation of educator, school, and district performance. State assessment data typically help inform the state's school and district accountability system, including assigning performance ratings to schools and districts. State assessment results are also typically a component of educator evaluation. CMAS is a source of data that

- may be used as a prompt for further investigation at the student, classroom, school, and district levels;
- supports districts/schools in reviewing and developing goals for the performance of their students, including subgroups;
- may indicate that a review of programs, curricula, materials, and/or scope and sequence may be appropriate; and
- may inform the evaluation of district/school approaches.

Assessment results also support a range of data-driven stakeholder conversations, activities, and decisions, including 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, to plan for curriculum development, and to report progress to parents. The results can also be used as one factor in making administrative decisions about program effectiveness, teacher effectiveness, class grouping, and needs assessment. CMAS results can also be used for research purposes and for informing community and organization efforts.

1.5. Assessment Development Partners

Activities specific to the CMAS assessments were conducted collaboratively by CDE, the Colorado educator community, and Pearson, the assessment contractor. Input and advice were provided by the Colorado Technical Advisory Committee (TAC), as well as by Achieve regarding the new science content standards and assessments.

1.5.1. Colorado Department of Education

As the administrative arm of the State Board of Education, CDE is responsible for implementing state and federal education laws. CDE's Assessment Unit works closely with Colorado school districts, educators, community stakeholders, and assessment development partners to develop and administer the state assessments. CDE focuses on creating assessments that serve students, schools, districts, and the community while complying with state and federal legal requirements. CDE content, assessment administration, special populations, technology, data, and psychometric staff work closely with Pearson on each facet of the assessment, with CDE serving as the ultimate approver of services and products provided.

1.5.2. Colorado Educator Community

Educator participation in the CMAS 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 potential bias and sensitivity issues. Throughout the test development process, educators participate in the following development activities, as shown in Table 1.1:

- Item writing: After receiving item writing assignments based on the CAS, educators create assessment items. Items that successfully move through the entire item development process will eventually appear on the operational assessments.
- Content and bias review: Educators review items to ensure content alignment and identify potential bias and sensitivity concerns before items are field tested.
- Rangefinding: Educators review student responses to field tested constructed-response items and define the score point ranges for the scoring rubrics that are used to score student responses.
- Data review: Before field tested items are included on operational assessments, educators review items with statistical parameters outside of normal ranges to determine if the item is acceptable for inclusion in the operational item bank.

Educator involvement in the new science assessment included the assessment development meetings included in Table 1.1 as well as the following:

- Review of blueprint and reporting structure: Educators reviewed different reporting structures, providing recommendations on how to achieve appropriate depth and breadth to the CMAS assessment.
- Unpacking the standards: Educators reviewed and revised unpacked standards that serve as a guide in item writing.
- Review of cognitive complexity matrix: Educators reviewed and revised the cognitive complexity matrix that will be used for the stimuli and items moving forward.
- Cognitive complexity review: A separate meeting was held to apply the revised cognitive complexity matrix to all newly developed stimuli and items.

Table 1.1. Schedule of Major Events

Event	Date(s)
ELA Passage Review	January 5 – March 12, 2021
ELA IWW Training	February 25, 2021
Math IWW Training	November 30, 2021
Content and Bias Review (Science)	February 22–25, 2021; March 2–4, 2021; September 13–16, 2021; November 8–12 and 18, 2021
Science SIMS Review	April 9, 2021
Content and Bias Review (Math)	May 17–21, 2021
Content and Bias Review (ELA)	July 26–30, 2021
DAC Administration Training	November 15–26, 2021
Spring 2022 Administration Window	April 11–29, 2022
Rangefinding (Science)	May 3–6, 2022; June 13–16, 2022
Rangefinding (Math, ELA)	June 6–24, 2022
Data Review (ELA)	August 18–19, 2022

Event	Date(s)
Data Review (Science)	August 23–24, 2022
Data Review (CSLA)	September 8, 2022
Data Review (Math)	January 18, 2023
Reports Available (ELA/CSLA/Math)	July 25, 2022
Reports Available (Science)	August 22, 2022
CMAS Science Standard Setting	September 27–28, 2022

1.5.3. Pearson

As the primary contractor responsible for the end-to-end assessment cycle services and products, Pearson works closely with CDE throughout the CMAS (all content areas) and CoAlt (science) assessment development and administration processes. This includes item and test development, forms creation, enrollment, packaging and distribution, test delivery, scoring, customer service, standard setting, scoring, score reporting, and psychometric services.

1.5.4. Tri-Lin Integrated Services, Inc.

As a subcontractor to Pearson, Tri-Lin is responsible for CSLA content and test development, including passage development, item development, and test form construction.

1.5.5. Measurement Inc.

As a subcontractor to Pearson, Measurement Inc. was responsible for CMAS Mathematics and ELA content and test development, including passage development, item development, and test form construction.

1.5.6. Colorado Technical Advisory Committee

The Colorado TAC is comprised of psychometric, assessment, and special populations experts tasked with providing high-level consulting and expert advice regarding validity and reliability issues. Topics for which the TAC has provided input include blueprint design, scaling and equating, mode comparability, scoring, reporting, and standard setting. The TAC included the following members during the 2022 assessment cycle:

- Dr. Jamal Abedi, Professor, University of California, Davis
- 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

1.5.7. Achieve

Pearson worked closely with Achieve during the development of the new science assessment. Achieve is an independent nonprofit education organization that leads the effort to help states make college and career readiness a priority for all students. Achieve provided background on how other states were approaching the new three-dimensional science standards and assessments, advice on how to proceed with cognitive complexity, blueprints, and reporting, and staff to co-facilitate a reporting and blueprint meeting with Colorado educators.

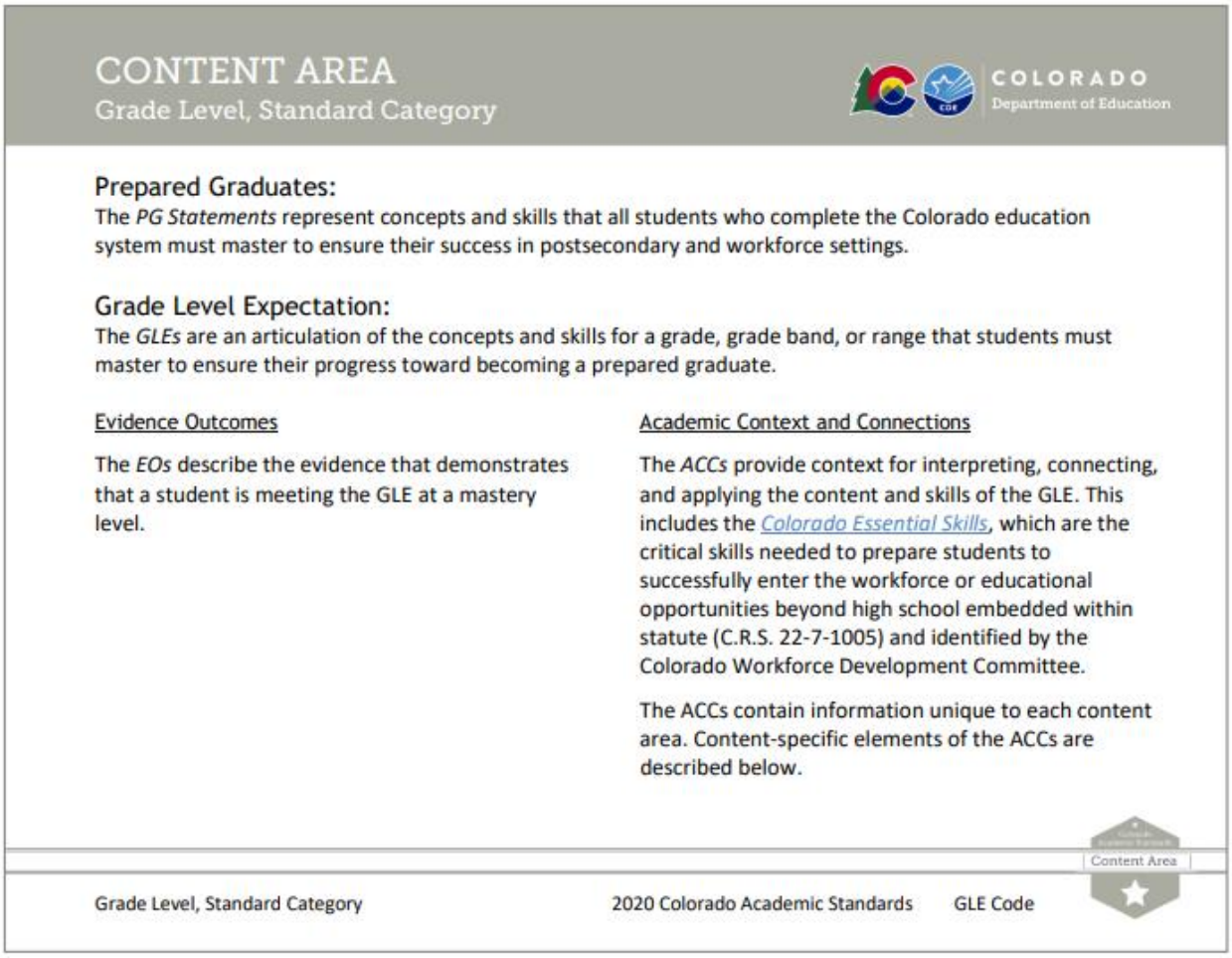
Chapter 2: Test Design

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.[3] The NGSS were guided by *A Framework for K–12 Science Education* (National Research Council, 2012). They are designed to reflect more recent research and thinking in science education and to better prepare students with the science knowledge, skills, and habits of mind to be ready for college, career, and civic responsibilities. As such, the 2020 CAS for Science represent what all Colorado students should know and be able to do in science because of their PreK–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[4]:

- 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:

[3]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.
[4]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 CMAS Science 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 percent representation and number of score points for each subclaim and standard area that appear on the assessments and 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 mathematics and ELA blueprints were developed in 2017–2018, while new test blueprints were created for the CMAS Science assessments in 2021–2022.

2.2.1. Mathematics and ELA

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 mathematics and ELA 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.[5]

2.2.2. Science

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 English learners. The blueprints were reviewed on October 14, 2021, by the TAC.

2.3. Claims and Subclaims

Student performance on the CMAS assessments is reported at the overall content area level as a scale score and accompanying performance level. Their performance is broken down even further to show performance at the claim and subclaim levels as shown in the test blueprints. The subclaims for the Mathematics assessments 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 the ELA assessments 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 Spring 2022 CMAS Science assessment reported students' percentile ranks with a comparison to other groups of students. More comprehensive reports with scale scores and performance level classifications will be provided for the new science assessment beginning in Spring 2023.

[5] 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.

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 ar guments, critiquing the reasoning of others and/or attending to pre cision 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 grad e/course by applying knowledge and skills articulated in the stand ards for the current grade/course (or for more complex problems, knowledge and skills articulated in the standards for previous grad es/courses), engaging particularly in the Modeling practice, and w here helpful making sense of problems and persevering to solve th em, reasoning abstractly and quantitatively, using appropriate tool s strategically, looking for the making use of structure, and/or loo king 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.
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. Performance Level Descriptors

Student performance on CMAS mathematics and ELA assessments is categorized into one of five performance levels: *Did Not Yet Meet Expectations, Partially Met Expectations, Approached Expectations, Met Expectations*, and *Exceeded Expectations*.[6] The performance

[6] The Spring 2022 CMAS Science assessment reported percentile ranks only. The new CMAS Science assessment will report scale scores and performance levels beginning with the Spring 2023 administration.

levels are based on the overall scale score. Cut scores divide the score scale for a grade and content area into five levels, as shown in Table 2.2. Students in the *Met Expectations* and *Exceeded Expectations* levels are considered on track to being college and career ready. 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: (1) 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 2.2, and (2) grade-level PLDs that provide detailed descriptions of performance levels by grade level and content area. The grade-level PLDs are available online at https://www.cde.state.co.us/assessment/cmas_plds and are included on the Individual Student Performance Report and in the *CMAS and CoAlt Interpretive Guide to Assessment Reports*.

Table 2.2. Performance Levels and Policy Claims—Mathematics and ELA Grades 3–8

Performance Level	*Did Not Yet Meet Expectations*	*Partially Met Expectations*	*Approached Expectations*	*Met Expectations*	*Exceeded Expectations*
Policy Performance Level Descriptor (PLD)	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

2.5. Cognitive Complexity

All mathematics and ELA items are tagged with a cognitive complexity level of high, moderate/medium, or low, as described in Table 2.3.

Table 2.3. 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 assessments now use a cognitive complexity matrix that examines items via three criteria, as shown in Table 2.4. Appendix B presents the most recent version of the matrix. Phenomenon in the stimulus material is examined separately for its own cognitive complexity.

Table 2.4. 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.

2.6. 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 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 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 in Appendix C and 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 only SR and PCR items.

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)

A subset of the science assessment includes item sets based on either interactive simulations or static information stimuli, which are groups of items that all relate to a scientific phenomenon, investigation, or experiment. Students use the information in the science simulations (SIMs), in the static cluster stimuli, and in the items to answer the items and make sense of phenomena. The items in these item sets may be SR, TE, or CR items.

2.7. Timing of Tests

Each assessment was composed of three sections with field test items embedded to allow the assessments to be administered in a reasonable timeframe, as shown in Table 2.5.

Table 2.5. Testing Times

Grades	Mathematics	ELA	Science
3–5	Sections 1–3: 65 minutes Total time: 195 minutes	Sections 1–3: 90 minutes Total time: 270 minutes	Sections 1–3: 80 minutes Total time: 240 minutes
6–8	Sections 1–3: 65 minutes Total time: 195 minutes	Sections 1–3: 120 minutes Total time: 360 minutes	Sections 1–3: 80 minutes Total time: 240 minutes
High School	N/A	N/A	Sections 1–3: 50 minutes Total time: 240 minutes

Chapter 3: Item Development

The item development process for the CMAS assessments involves following prescribed steps to develop 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 (IDP) 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.

As part of the test construction process, a selection of the proposed set of operational items are refreshed, as illustrated in Table 3.1. Therefore, a portion of the operational items had been used operationally on a previous CMAS form, while the remaining items are refreshed using Colorado-developed field test items. The Spring 2022 CMAS Science assessments included a set of core items held constant across all forms and embedded field test items differing from form to form. The core items were used as operational items for scoring purposes. All items were reviewed by Colorado educators.

Table 3.1. Refresh Rates—Mathematics and ELA

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%	–

3.1. Item Banking System

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 IDP 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 directly to the ES, EOs, and the corresponding CAS can be created.

3.3. ELA Passage Development

Item development for ELA begins with the research and selection of high-quality literary and informational texts. Due to the availability of appropriate passages and challenges with acquiring permissions, passages to be used on the CSLA forms are commissioned by Tri-Lin, either in-house or by professional passage writers. 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 IDP. The passage selection (and writing) guidelines, task model descriptions, and cognitive complexity framework define the number of texts/passages by text type, genre, length, and complexity. 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 the passages are accepted by CDE, committees of educators review them for content and bias. The committees are comprised of educators from throughout the state representing a variety of student populations, including students with disabilities and students with limited English proficiency. Passages accepted by both CDE and the educator committees are then used for item writing.

3.4. Science Simulation and Cluster Stimulus Development

Item development for science begins with the preliminary conception and composition of the science SIMs and cluster stimuli. The number and types of needed SIMs and cluster stimuli are determined by the test construction specifications, a gap analysis of the pool of available SIMs and stimuli, and the IDP. Topics for SIMs and cluster stimuli are researched for suitability of science content, alignment to the standards, and grade-level appropriateness.

SIMs and cluster stimuli follow slightly different paths through the development process, but both include multiple steps of review and evaluation by assessment specialists using approved criteria, including 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.

SIM ideas are presented to CDE for review and feedback in the form of storyboards illustrating the intended virtual interaction, along with suggested EOs that the SIMs address. CDE provides feedback on how to move forward with the development of the SIMs. Revised storyboards are then reviewed by committees of educators from throughout the state representing a variety of student populations, including students with disabilities and students with limited English proficiency. The SIMs are then fully developed into animated interactions and again 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 as used for the SIMs.

3.5. Item Writing

Item writer workshops (IWWs) with Colorado educators were conducted for the 2021–2022 development cycle for mathematics and ELA, whereas item development for the Spring 2022 CMAS Science assessment was conducted by professional item writers who participated in contractor item writer trainings in Fall 2020 and Winter 2021.

IWW participants are 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 was needed for CSLA items. Content specialists from CDE 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 authored the items in ABBI, where Pearson or Tri-Lin assessment specialists completed their initial review. The assessment specialists reviewed and suggested revisions to the items and metadata for the item authors, who then made the revisions and resubmitted 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.

When appropriate, human-scored CR items are reviewed for their scorability by a performance scoring director, and items and/or scoring rubrics with score points deemed "difficult to score" are revised in collaboration with the assessment specialist(s). Equation editor/CR items scored by an automated engine are reviewed for their scorability by a mathematics product analyst, and items and/or scoring rubrics are revised in collaboration with the assessment specialist(s).

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, socio-economic 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 completed, 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 assessment specialists. CDE's comments and determinations regarding the status of the items, as indicated below, are recorded in ABBI:

- 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 Pearson's 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. The purposes of this educator review are to (1) ensure that the items are properly aligned to the CAS, accurately measure the intended content, and are grade-appropriate; and (2) 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 Colorado educators are selected 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 English learners). 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 ELs, 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 committee members are trained and instructed to verify that each item and stimulus (list non-exhaustive):

- displays and functions correctly in TestNav 8 Preview (i.e., Pearson's online testing platform students use to access the assessment);
- aligns to the ES and/or EO;
- uses clear, unambiguous, and grade-level appropriate language;
- avoids construct-irrelevant complex sentence structure;
- uses everyday words to convey meaning when vocabulary is not part of the tested construct;
- has one correct answer (depending on the item type);
- contains plausible distractors that represented feasible misunderstandings of the content (depending on the item type);
- represents the range of cognitive complexities and included 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 committee 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 SIMs/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 if the functioning of the items supports their inclusion in the item pool used for operational forms construction. Following the administration of items in a field test environment, psychometricians perform statistical analyses on the field-tested items that are used to evaluate their quality. Table 3.2 presents the statistical flags applied to the field test items.

Table 3.2. 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% zeros, the item is flagged. The rule is 50%+ zeros for mathematics, ELA, and CSLA and 60%+ for science.
***B DIF indicates moderate DIF, whereas C DIF indicates significant DIF.

During data review, a committee of educators reviews the flagged items and their statistics along with student performance data. Separate data review committees are convened for each content area, including the accommodated CSLA items. Participants are provided item images and metadata, along with classical and DIF statistics.

Classical statistics included item means, item-total correlations/point biserials, and distribution of responses across answer options or score points, depending on item type. Items were flagged based on the criteria in Table 3.2, and flagged items were taken to data review.

DIF analyses for CMAS items were conducted on various subgroups (gender, ethnicity, free and reduced lunch, IEP, and ELs) using Mantel–Haenszel Delta DIF statistics (Dorans & Holland, 1992). The same analysis methods were used for CSLA items, but the DIF analyses were 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 that are classified as moderate or significant DIF are taken to data review.

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 3.3 presents the final results following the data review based on Spring 2022 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 3.3. Data Review Results

Content Area	Grade	#Accepted	#Accepted for Revision and Re-field test	#Rejected
Mathematics	3	14	0	2
	4	6	0	1
	5	15	0	2
	6	24	0	4
	7	21	0	2
	8	26	0	1
ELA	3	9	0	1
	4	26	0	4
	5	32	0	1
	6	30	0	2
	7	36	0	1
	8	78	0	5
CSLA	3	3	0	3
	4	4	0	2
Science	5	182	0	21
	8	199	3	10
	11	247	12	35

Chapter 4: Test Construction

The Spring 2022 CMAS Science test forms were brand-new with items administered to Colorado students for the first time, whereas the Spring 2022 mathematics and ELA test forms were either test forms intended for use in the postponed 2019–2020 test administration or were newly developed test forms. New operational test forms were administered in Spring 2022 for ELA Grades 3, 5, and 7 and mathematics Grades 4, 6, and 8, whereas test forms intended for use in 2019–2020 were administered for ELA Grades 4, 6, and 8; mathematics Grades 3, 5, and 7; and CSLA. The test forms were constructed through an iterative process between Pearson and Measurement Inc. staff for CMAS mathematics and ELA and Pearson and Tri-Lin staff for CSLA. Pearson was solely responsible for construction of the science test forms. Once the test forms were constructed, CDE reviewed the forms, provided feedback, and gave final approval.

4.1. Test Form Construction Process

When building the test forms, Pearson, Measurement Inc., or Tri-Lin 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. The following guidelines for were used during the Spring 2022 form construction:

- Adherence to the test blueprints and test construction specification targets
    - Exact match to blueprint for subclaims
    - Same distribution of cognitive complexity
    - Same percentage of TEs
- Review of the item statistics and adherence to the statistical criteria in the test construction specifications (mathematics and ELA only[7])
    - Evaluation of item means, point biserial correlations, and score point distributions
    - Evaluation of IRT item parameter estimates
    - Evaluation of item fit statistics
    - Mirroring of 2018 test characteristic curves and conditional standard error of measurement (CSEM) curves (mathematics and ELA only[8])
- 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 instances of clueing and/or content overlap

[7] Statistics were not used to inform the Spring 2022 test build for science as it was a new assessment.
[8] There was no mirroring of the CSEM curves for the Spring 2022 science assessment as the items did not have statistics prior to field testing.

After the initial operational item pull is complete, assessment specialists verify that the test forms meet 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.[9] (See Chapter 9: 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.

After the operational form is approved, field test items are selected from the item bank. 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. For CMAS Science, a set of items designated as 'operational' was included on all forms and used for producing student raw scores. The items that varied across forms functioned as an embedded field test and were not included in scoring.

4.2. Online Forms

Most students take the CMAS assessments online. Using this format allows not only for the use of innovative item types but also for additional accessibility options and accommodations as described in Chapter 5: Test Administration (e.g., text-to-speech).

4.3. Accommodated Test Forms

The online testing format allows 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. Accommodations are also available for students who need them and include paper, large print, braille forms, and oral scripts, as well as online forms designed to work with assistive technology such as screen readers. For paper forms, the various options are described below. Oral scripts are also available for the online and paper forms in both English and Spanish for mathematics and science. English oral scripts are available for local translation into languages other than Spanish. Due to the effort involved in creating an approved accommodated form, these forms are not refreshed at the same rate as the online forms. For the Spring 2022 administration, the mathematics assessments designed to work with assistive technology at grades 4, 6, and 8 used the same form intended for use in 2019–2020. The reuse administrations for the paper and braille forms are provided below.

4.3.1. Paper

Paper-based versions of the CMAS assessments are available if needed for an accommodation or for schools that choose not to test online as allowed by state law. A Spanish transadaptation is also available on paper for CMAS Mathematics and Science. CSLA is the accommodated version of CMAS ELA for eligible Spanish-speaking students in Grades 3 and 4 and is administered on paper.

[9] This step in the form construction process was not applicable to the Spring 2022 CMAS Science assessment.

The paper form is parallel to the online form, meaning the paper and online forms include 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. In some cases, this was achieved with traditional SR items, and in others it 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 statistics of the TE item are compared to the paper-based version to confirm equivalence.

For the spring 2022 administration, the operational items on the ELA and science paper-based forms were the same as the operational items on the online forms. For the mathematics assessments at grades 4, 6, and 8, the English version of the paper-based form was the same as the operational items on the English version of the online form. The Spanish version of the paper-based form was the same as the operational items intended for use in 2019–2020.

4.3.2. Braille

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.

For the Spring 2022 administration, the mathematics assessments at grades 4, 6, and 8 used the same braille form intended for use in 2019–2020.

4.3.3. Large Print

Large print versions of the CMAS assessments are also created. The large print versions 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 that students use to access the assessment. PearsonAccess^{next} is the student test management portal that 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 and in-person and recorded trainings.

5.1. Manuals

The following manuals are 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 that 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.
- The *CMAS and CoAlt Procedures Manual* provides instructions for the coordination of the CMAS assessments. Instructions include the protocols that all school staff are to follow related to test security, test administration, and providing accommodations to students with disabilities and English learners 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 to ensure the comparability of resulting scores and accurate interpretation of results.

Live virtual trainings were conducted by CDE for groups of DACs based on their level of experience in the position. During the training, 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 there was an opportunity for questions and answers. Resources used during the live trainings are posted on the CDE website at http://www.cde.state.co.us/assessment/trainings-archive. Administration training materials such as slide decks, manuals, and how-to guides were also

available on the CDE Assessment Unit website for training SACs and Test Administrators. 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 thoroughly and knowledgeably about the administration and to escalate inquiries as necessary. A knowledge base of commonly asked questions was created to ensure accurate and consistent responses to school and district personnel. The knowledge base was created by CDE and Pearson based on information covered in the training materials and manuals. Revisions and additions were made to the knowledge base 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.

5.3. Practice Resources

Colorado Practice Resources (CPRs) are available online at https://coassessments.com/practice-resources/ to help students become familiar with the item types on the CMAS assessments. The CPRs are continually updated as needed to reflect current accessibility features and any updates to TestNav that may impact student interactions with the system. Accommodated versions of the CPRs are also available so that students can practice using accommodations and accessibility features such as text-to-speech, color contrast, and Spanish text-to-speech. Paper sample items for students taking the paper versions of the assessments (including CSLA) are available in PDF format for download and 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 with tools and resources to test their environment readiness status. Issues identified from site readiness evaluations were assessed by Pearson and CDE, and appropriate corrective actions were developed and communicated to affected districts.

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 and 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 the text-to-speech accessibility feature is available to all students, only students who need text-to-speech are assigned to it in advance of testing. Similarly, the CSLA assessments were developed to be linguistically accommodated Spanish tests and, as such, is designed to be linguistically accessible for eligible Spanish-speaking students.

Accommodations are also available to the population of students with IEP or 504 plans or students who are English learners. For example, students may have extended time as required by their IEP or as allowed for students classified as English learners. 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 a student with an opportunity to access 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. The accommodations of calculator on non-calculator sections of mathematics and scribe for CR items on ELA and 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 oral scripts (mathematics and science), Spanish oral scripts (mathematics and science), oral 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, and Spanish forms with and without text-to-speech for mathematics and science.

Live webinar accommodations and accessibility features training was conducted by CDE for district-level personnel. The intent of this training was 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. A recorded version of the live training, slide decks, and procedural information (Section 6.0 of the *CMAS and CoAlt Procedures Manual*) are available on the CDE website at http://www.cde.state.co.us/assessment/trainings-archive.

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 computer functions not necessary to complete the test 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 at the student level, 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.

Chapter 6: Scoring

The CMAS assessments use a combination of machine and human scoring. The SR and TE items are machine-scored, with point values varying by item type and assessment. The CMAS Mathematics and Science CR items are hand scored. The CMAS ELA PCR items are scored on two trait dimensions using a combination of human scoring and automated scoring. Pearson's Performance Scoring team implemented the scoring process for CR, human PCR, and parallel paper-based versions of the TE items for CMAS. The holistic rubrics used to score both the CMAS ELA and CSLA PCR items can be found in Appendix C.

To maintain comparability with the scoring prior to 2022, scoring rules for the SR and TE machine-scored items and CR items were preserved from previous years. CR items used prior years' rubrics, anchor papers, rules and scoring methods to maintain comparability. An exception for this is the Spring 2022 CMAS Science assessments that had no historical knowledge for comparison.

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. Human Scoring

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 (for all content areas except Science[10]). 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. With distributed scoring, scorers are able to score seven days per 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 only completed 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

[10] Science leveraged baseline qualification for most items; scorers qualified on an item they did not score before proceeding to item-level training that did not contain qualification. Only four of the science operational items had qualification sets developed due to the nature of the assessment where field test items were in operational slots.

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 is 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 and CSLA use holistic rubrics for each item type (as shown in Appendix C). 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. 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 are continued to be spot-checked as an added confirmation of their accuracy. The agreement rate requirements are as follows:

- 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

Section 11.5 of this technical report presents an explanation of the rater agreement statistics and the results from the Spring 2022 administration.

6.2.5. Calibration

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. Automated Scoring

Pearson's Intelligent Essay Assessor (IEA) was used for scoring three out of the 12 operational PCRs in 2022 for the CMAS ELA assessment. Two had an automated scoring model based on training from prior operational years, one was trained based on 2018 field test data, and the remaining nine were scored by human scorers. Items that used automated scoring were also checked for quality using second scores by human scorers. Ten percent of responses were randomly selected and given a second reliability score to provide data for evaluating the consistency of scoring. Some responses were not scored by the engine at all and received a first human score based on Smart Routing.

6.3.1. Smart Routing

The use of smart routing during operational scoring increases the quality of automated scoring by routing responses that are more likely to disagree with a human score to receive an additional human score. When human scorers read a response, they typically apply integer scores based on a scoring rubric. For example, when there is strong agreement between two independent human scorers, they might both assign a score of 3, such that the average score over both raters is also a 3 (i.e., (3+3)/2 = 3). IEA simulates this behavior, but because its scores come from an artificial intelligence algorithm, it generates continuous (i.e., decimal-valued) scores. In the previous example, the IEA score might be a 2.9 or 3.1. Similarly, if the human scorers disagreed on a response and scored it as a 3 and a 4, IEA would likely provide a score between 3 and 4 (e.g., 3.4 or 3.6). This continuous IEA score needs to be rounded to an integer score for reporting (i.e., a 3 or a 4, depending on rounding rules, in this example).

Smart routing involves routing the responses where the IEA score tends to disagree with human scores for additional human review. Because the cases that result from "in between" scores are based on modeling human scores, it follows that human scores may be less certain as well. Therefore, responses are more likely to be double-scored and resolved if the IEA and human scores are non-adjacent. Smart routing is used as needed to achieve targeted quality metrics (e.g., validity agreement or agreement with human scorers). Smart routing involves the application of the following steps:

1. The continuous IEA score for each of the two trait scores is rounded to the nearest score interval of 0.2, starting from 0. For example, IEA scores between 0 and 0.1 are rounded to an interval score of 0, scores between 0.1 and 0.3 are rounded to an interval score of 0.2, scores between 0.3 and 0.5 are rounded to an interval score of 0.4, etc.
2. Within each of these intervals, the percentage of exact agreement between IEA integer scores and the human scores is calculated for each trait.
3. For each prompt, agreement rates are evaluated for each interval and for each trait.
4. Responses within intervals for which IEA–human agreement on either trait are below a designated threshold are routed for additional human scoring.

6.3.2. Quality Criteria for Evaluating Automated Scoring

The primary evaluation criteria for IEA are based on responses to validity papers with "known" scores assigned by experts. For each prompt scored, a set of validity papers is used to monitor the human-scoring process over time. Validity papers are seeded into human scoring throughout the administration. The expectation is that IEA can score validity papers at least as accurately as humans can score the papers.

Additional measures of inter-rater agreement for evaluating automated scoring are proposed based on the research literature (Williamson et al., 2012) and include Pearson correlation, kappa, quadratic-weighted kappa, exact agreement, and standardized mean difference. These measures are computed between pairs of human scores, as well as between IEA and humans, to evaluate how performance is the same or different. Criteria for evaluating the training of IEA given these measures include the following:

- Pearson correlation between IEA and human scores should be within 0.1 of human–human correlation.
- Kappa between IEA and human scores should be within 0.1 of human–human kappa.
- Quadratic-weighted kappa between IEA and human scores should be within 0.1 of human–human quadratic-weighted kappa.
- Exact agreement rate for IEA and human scores should be within 3.0% of the human–human exact agreement rate.
- Standardized mean difference (SMD) between IEA and human scores should be less than 0.15.

The specific criteria for evaluating IEA included both primary and secondary criteria, as described below.

6.3.2.1. Primary Criterion

The performance of IEA was evaluated by comparing IEA scores with human scores for the set of validity papers. The primary criterion is stated as follows: *With smart routing applied as needed, IEA agreement is as good as or better than human agreement for each trait score.* For a given prompt, this criterion is operationalized as follows:

1. Determine agreement of the human scores with the validity papers for each trait.
2. Calculate agreement of the IEA scores with the validity papers for each trait.
3. Compare the IEA-human agreement on the validity papers.
4. Deploy IEA operationally if the IEA validity agreement is greater than or equal to the human agreement for each trait.

When it is not possible to use human-scored validity responses in evaluating IEA performance, IEA is evaluated based on IEA–human exact agreement for each trait score and compared to agreement based on responses that are double-scored by humans. IEA-human agreement is evaluated on a portion of the data according to the following steps:

1. Determine exact agreement of the two human scores with each other for each trait.
2. Calculate agreement of the IEA scores with the human scores for each trait.
3. Compare the IEA-human agreement with the human-human agreement.
4. Deploy IEA operationally if the IEA-human agreement is within 5.25% of the human-human agreement.

In addition to the overall comparison, the following performance thresholds are targeted in the test dataset: (1) at least 65% overall IEA-human agreement and (2) 50% IEA-human agreement by score point (i.e., conditioned on the human score). These targets go beyond the contingent primary criteria approved by the consortium state leads.

6.3.2.2. Secondary Criterion

The secondary criterion involves comparing agreement indices for IEA-human scoring for various demographic subgroups and is stated as follows: *With smart routing applied as needed, IEA-human differences on statistical measures for each trait score are within the Williamson et al. (2012) tolerances for subgroups with at least 50 responses.* IEA-human agreement is evaluated according to the following steps:

1. Determine exact agreement of the two human scores with each other for each trait.
1. Calculate agreement of the IEA scores with the human scores for each trait.
2. Compare the IEA-human agreement with the human–human agreement.
3. For subgroups with at least 50 IEA-human scores and at least 50 human-human scores, compare agreement indices to the following criteria:
    a. Pearson correlation between IEA-human should be within 0.1 of human-human.
    b. Kappa between IEA-human should be within 0.1 of human–human.
    c. Quadratic-weighted kappa between IEA-human should be within 0.1 of human-human.
    d. Exact agreement between IEA-human should be within 5.25% of human-human.
    e. SMD between IEA-human should be less than ±0.15 (this criterion was applied to subgroups with at least 50 IEA-human scores).

Although it is not expected that these criteria will be met for all subgroups for all prompts, if results of the evaluation between IEA and human scoring for subgroups for any prompt indicate that IEA performance persistently fails on the criteria listed above, considerations would be given to resetting the responses scored by IEA and reverting to human scoring until such time that an alternate IEA model could be established with improved subgroup performance.

In addition to the secondary criterion, the performance of IEA is also compared with the following targets on the various measures for subgroups with at least 50 responses:

- Pearson correlation between IEA-human should be 0.70 or above.
- Kappa between IEA-human should be 0.40 or above.
- Quadratic-weighted kappa between IEA-human should be 0.70 or above.
- Exact agreement between IEA-human should be 65% or above.

6.3.3. Hierarchy of Assigned Scores for Reporting

When multiple scores are assigned for a given response, the following hierarchy determines which score is reported operationally:

- The IEA score is reported if it is the only score assigned.
- If an IEA score and a human score are assigned, the human score is reported.
- If two human scores are assigned, the first human score is reported.
- If a backread score and human and/or IEA scores are assigned, the last backread score is reported.
- If a resolution score is assigned, the resolution score is reported. If nonadjacent scores are encountered, responses are automatically routed to resolution.

Chapter 7: Standard Setting

To support the interpretation of student results, student performance on the CMAS assessments is described in terms of five performance levels as presented in Table 2.2. Standard setting is the process of translating those policy-driven performance standards into scores on the assessment. The purpose of a standard setting study 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 7.1 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). Details of the standard setting process can be found 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 in Fall 2022 using the spring 2022 data; the standard setting report is still in progress, and details of the meeting and the final cuts will be included in the 2022–2023 technical report.

Table 7.1. Performance Level Cut Scores

Content Area	Grade	*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
	4	650–699	700–724	725–749	750–795	796–850
	5	650–699	700–724	725–749	750–789	790–850
	6	650–699	700–724	725–749	750–787	788–850
	7	650–699	700–724	725–749	750–785	786–850
	8	650–699	700–724	725–749	750–800	801–850
ELA	3	650–699	700–724	725–749	750–809	810–850
	4	650–699	700–724	725–749	750–789	790–850
	5	650–699	700–724	725–749	750–798	799–850
	6	650–699	700–724	725–749	750–789	790–850
	7	650–699	700–724	725–749	750–784	785–850
	8	650–699	700–724	725–749	750–793	794–850
CSLA	3	650–699	700–724	725–749	750–778	779–850
	4	650–699	700–724	725–749	750–771	772–850

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 7.2. This cut score 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 7.2. ELA Reading *Met Expectations* Cut Score

Scale Range	Cut Score
110–190	150

Chapter 8: Reporting

Several score reports are generated to communicate student performance on the CMAS assessments. The reports contain a variety of score types at different levels of the blueprint, as described in this section. For additional details on score reports, see the *CMAS and CoAlt Interpretive Guide to Assessment Reports*.

Note: Because Colorado students saw CMAS and CoAlt Science items for the first time in spring 2022, standard setting was held after the spring 2022 reporting cycle. Science reporting for spring 2022 assessments provided normative achievement indicators in the form of percentile ranks but did not include criterion referenced reporting.

8.1. Description of Scores

CMAS reports provide information on student performance in terms of scale scores, performance levels, percentile ranks, and percent earned scores. Because Colorado students saw new science content for the first time in spring 2022 and a standard setting had not been held, scale scores and performance levels were not reported for science.

8.1.1. 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. Students taking CMAS mathematics and ELA assessments receive scale scores in each of the following areas: (1) overall test and (2) Reading claim (ELA and CSLA only). The overall scale for each test ranges from 650 to 850, and the ELA Reading scale ranges from 110 to 190, as shown in Table 7.1 and Table 7.2 in the previous chapter.

8.1.2. Performance Levels

Performance levels and their accompanying PLDs are reported at the overall assessment level. Students are classified into performance levels based on their scale score and the cut scores obtained from standard setting as described in the previous chapter. 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*. Students in the top two categories (i.e., *Met Expectations* and *Exceeded Expectations*) are considered to be on track to being college- and career-ready in that content area.

8.1.3. 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. For CMAS Mathematics and ELA, the percentile rankings were based on the overall scale score, but for the Spring 2022 CMAS Science assessments, percentile rankings were based on the overall raw score.

8.1.4. Percent Earned

To prevent incorrect interpretations and provide a metric that is more generally understood, students' performance for the Writing claim (ELA and CSLA) and the subclaims in ELA, CSLA, and mathematics are reported as the percentage of points earned.

The percent of points earned refers to the number of points a student earned out of the total number of points possible within a claim or subclaim. Unlike scale scores, 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. In addition, performance on different subclaims or subscales 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 score 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.

8.2. Score Reports

Two types of score reports are provided: (1) the student-level Student Performance Report and (2) the aggregate reports at the school and district levels. For a detailed explanation of the information provided in the reports, refer to the *CMAS and CoAlt Interpretive Guide to Assessment Reports*. Appendix D presents sample student performance reports. CSLA assessments are parallel and comparable to CMAS ELA assessments in scoring and reporting. Therefore, separate CSLA reports are not included (please refer to the CMAS ELA examples).

8.2.1. Student Performance Reports

The Student Performance Report provides information about the performance of a particular student. The student's scale score(s), associated 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. [11] 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.

8.2.2. Aggregate Reports

Several types of aggregate reports are produced for schools and districts:

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

These reports are produced at the school and/or district levels and provide summary information for a given school or district. District and school reports are provided electronically through PearsonAccess^{Next}. 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.

[11] The Spring 2022 CMAS Science Student Performance Report included a student's percentile rank only as compared to the rest of the state.

Access to the reports is limited to authorized users. Examples of each type of aggregate report and a detailed explanation are provided in the *CMAS and CoAlt Interpretive Guide to Assessment Reports*.

Chapter 9: Test Results and Analysis

This chapter presents the test results and statistical analyses for the Spring 2022 CMAS assessments in mathematics, ELA (including CSLA), and science.

9.1. Student Participation

Table 9.1 presents a breakdown of the number of students who took the assessment online compared with those who took accommodated forms. Although a paper form was available to all students, most students took the assessments online. Appendix E presents n-counts for various demographic characteristics for the students who took the CMAS assessments.

Table 9.1. Student Participation N-Count by Form

Content Area	Form	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
Mathematics	Online	39,819	41,461	43,272	47,352	47,308	45,806	–
	Spanish Online	1,062	937	518	235	253	226	–
	Paper	4,489	4,193	4,035	3,751	3,630	3,455	–
	Spanish Paper	122	59	32	7	6	10	–
	Text-to-Speech	10,987	10,226	9,562	4,579	4,078	3,298	–
	Assistive Technology	3	10	4	12	8	9	–
	Total	56,482	56,886	57,423	55,936	55,283	52,804	–
ELA	Online	50,399	51,416	53,233	52,513	51,993	49,764	–
	Paper	4,675	4,315	4,124	3,424	3,258	2,951	–
	Text-to-Speech	0	0	0	1	0	0	–
	Assistive Technology	7	15	5	22	14	12	–
	Total	55,081	55,746	57,362	55,960	55,265	52,727	–
CSLA	Total	1,301	1,151	–	–	–	–	–
Science	Online	–	–	43,792	–	–	45,518	28,740
	Spanish Online	–	–	170	–	–	158	79
	Paper	–	–	3,082	–	–	2,467	971
	Spanish Paper	–	–	2	–	–	10	2
	Text-to-Speech	–	–	9,568	–	–	3,300	777
	Spanish Text-to-Speech	–	–	206	–	–	74	55
	Total	–	–	56,820	–	–	51,527	30,624

9.2. Performance Results

Appendix F presents the cumulative scale score distributions by grade, and Appendix G displays the same information in graphical form. Based on the review of adjudication results, one item on the CMAS Mathematics Grade 6 accommodated assessment and two items on the CMAS Science Grade 8 assessment were removed and suppressed from scoring.

Table 9.2 presents summary statistics for overall scale scores, including the means, standard deviation (SD), and median scale scores. The previous year's results are also included for comparison. Appendix H presents the summary statistics for the overall scale scores by demographic subgroup. The tables also include the coefficient alpha (see Section 11.1 for more information). Table 9.3 presents the performance level distributions and includes the distributions from last year's administration for comparison, and Table 9.4 presents the percentile rank performance summary for science.

Table 9.2. Scale Score Performance Summary

Content Area	Grade	2022				2021			
		N	Mean	SD	Median	N	Mean	SD	Median
Mathematics	3	56,482	737	39.3	738	5,941	741	37.7	742
	4	56,886	732	32.9	732	46,782	729	33.5	729
	5	57,423	736	35.1	734	5,935	739	34.1	738
	6	55,936	728	32.5	728	44,817	727	32.4	726
	7	55,283	730	27.7	728	5,510	733	26.9	731
	8	52,804	731	40.1	728	39,202	730	38.8	726
ELA	3	55,081	737	43.9	739	45,191	736	42.6	737
	4	55,746	740	36.4	744	5,725	743	34.7	747
	5	57,362	745	32.8	745	46,917	746	32.4	747
	6	55,960	742	34.1	743	5,439	738	30.3	739
	7	55,265	741	37.2	742	42,934	742	37.0	742
	8	52,727	742	40.9	743	4,463	743	38.1	744
ELA – Reading	3	55,081	145	17.7	145	45,191	145	17.6	144
	4	55,746	146	14.6	148	5,725	149	14.8	150
	5	57,362	148	13.2	148	46,917	149	13.4	149
	6	55,960	147	13.7	147	5,439	147	12.9	147
	7	55,265	146	14.8	146	42,934	147	15.4	147
	8	52,727	147	16.2	147	4,463	149	16.0	149
CSLA	3	1,301	726	26.9	726	837	723	25.5	725
	4	1,151	726	22.0	727	53	733	26.3	741
CSLA – Reading	3	1,301	140	10.5	139	837	140	10.0	140
	4	1,151	141	8.3	140	53	143	10.3	147

Table 9.3. Performance Level Distribution: Percent of Students in Each Performance Level

Content Area	Grade	2022					2021				
		1	2	3	4	5	1	2	3	4	5
Mathematics	3	18.68	18.95	23.00	30.73	8.64	14.98	17.87	24.54	33.47	9.14
	4	18.05	24.05	27.20	28.28	2.43	20.09	24.94	26.45	26.24	2.27
	5	16.17	24.31	24.59	28.06	6.86	13.72	23.18	24.45	31.46	7.19
	6	20.33	26.02	27.37	22.84	3.44	21.03	27.11	27.73	20.94	3.20
	7	12.01	32.73	30.21	22.49	2.56	10.00	30.93	31.80	24.74	2.54
	8	23.60	23.05	20.97	27.42	4.96	22.49	25.49	22.49	24.80	4.73
ELA	3	22.14	16.56	20.55	35.86	4.88	21.54	17.95	21.37	34.90	4.23
	4	15.11	15.59	25.18	36.85	7.27	12.37	15.81	25.89	39.48	6.46
	5	8.21	19.34	27.05	40.62	4.79	7.66	19.21	25.94	42.56	4.63
	6	11.49	18.52	26.95	35.64	7.40	10.04	22.65	30.78	32.40	4.14
	7	15.09	18.93	24.21	29.46	12.30	13.33	19.08	24.96	29.64	12.99
	8	16.65	17.08	22.33	33.29	10.65	14.21	17.10	23.68	36.14	8.87
CSLA	3	18.45	27.98	33.74	17.60	2.23	15.77	33.81	35.01	14.46	0.96
	4	11.56	33.54	41.18	12.51	1.22	16.98	13.21	43.40	22.64	3.77

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.

Table 9.4. Science Percentile Rank Performance Summary

Content Area	Grade	N	Mean	SD	Median*
Science	5	56,820	50	28.8	51
	8	51,525	50	28.8	49
	11	30,622	50	28.8	49

* The median is not 50 where there are an even number of obtained scores being ranked.

Appendix I 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-correct 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

Appendix J presents the item-level classical statistics for each CMAS and CSLA assessment, including the omit rate, *p*-value, item-total correlation, and the percentage of students earning each score point (CR items only).

Item difficulty is measured by the *p*-value bounded by 0.0 and 1.0 that indicates how easy or hard an item is. The *p*-value for 1-point items is based on the proportion of students who answered an item correctly and is derived 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) bounded by -1.0 and 1.0 that 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.

9.4. Subclaim Correlations

The ELA and 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 subclaim scores: Physical Science, Life Science, Earth Systems Science, and Science and Engineering Practices (SEPs).

One way to assess the internal structure of a test is through the evaluation of correlations among subscores, as presented in Table 9.5–Table 9.9. For CMAS ELA and CSLA, these analyses were conducted between the Reading and Writing claim scores and the subclaims (RL, RI, RV, WE, and WKL). For CMAS Mathematics and Science, the analyses were conducted between the subclaim scores. There is evidence of unidimensionality if the components within a content area are strongly related to each other.

The intercorrelations for the mathematics subclaims were higher overall than the ELA and CSLA intercorrelations, although most values for all assessments were between 0.35 and 0.95. For CMAS ELA and CSLA, the two writing subclaims tended to have higher correlations with one another than they did with any of the reading subclaims.

Table 9.5. Correlations Between Subclaims—Mathematics

Grade	Subclaim	Subclaim B	Subclaim C	Subclaim D	Total Test
3	A	0.828	0.730	0.758	0.956
	B	–	0.689	0.725	0.899
	C	–	–	0.692	0.851
	D	–	–	–	0.862
4	A	0.735	0.787	0.723	0.959
	B	–	0.659	0.614	0.809
	C	–	–	0.693	0.891
	D	–	–	–	0.838
5	A	0.781	0.784	0.765	0.960
	B	–	0.700	0.711	0.865
	C	–	–	0.747	0.883
	D	–	–	–	0.873
6	A	0.700	0.756	0.742	0.939
	B	–	0.649	0.623	0.823
	C	–	–	0.725	0.887
	D	–	–	–	0.859

Grade	Subclaim	Subclaim B	Subclaim C	Subclaim D	Total Test
7	A	0.662	0.776	0.735	0.943
	B	–	0.634	0.597	0.783
	C	–	–	0.740	0.900
	D	–	–		0.858
8	A	0.703	0.783	0.759	0.954
	B	–	0.653	0.667	0.813
	C	–	–	0.737	0.884
	D	–	–	–	0.874

Table 9.6. Correlations between Subclaims—Reading vs. Writing

Content Area	Grade	Correlation
ELA	3	0.683
	4	0.681
	5	0.680
	6	0.730
	7	0.712
	8	0.741
CSLA	3	0.663
	4	0.674

Table 9.7. Correlations Between Subclaims—ELA

Grade	Subclaim	RI	RV	WE	WKL	Total Test
3	RL	0.696	0.652	0.661	0.557	0.889
	RI	–	0.695	0.553	0.492	0.862
	RV	–	–	0.525	0.464	0.815
	WE	–	–	–	0.654	0.821
	WKL	–	–	–	–	0.697
4	RL	0.730	0.676	0.678	0.618	0.899
	RI	–	0.654	0.574	0.555	0.866
	RV	–	–	0.514	0.474	0.781
	WE	–	–	–	0.817	0.847
	WKL	–	–	–	–	0.781
5	RL	0.657	0.573	0.570	0.539	0.852
	RI	–	0.537	0.659	0.601	0.859
	RV	–	–	0.452	0.432	0.697
	WE	–	–	–	0.810	0.860
	WKL	–	–	–	–	0.789
6	RL	0.686	0.589	0.640	0.618	0.863
	RI	–	0.625	0.691	0.647	0.882
	RV	–	–	0.509	0.499	0.727
	WE	–	–	–	0.875	0.884
	WKL	–	–	–	–	0.837

Grade	Subclaim	RI	RV	WE	WKL	Total Test
7	RL	0.718	0.676	0.718	0.693	0.898
	RI	–	0.629	0.606	0.595	0.849
	RV	–	–	0.531	0.524	0.763
	WE	–	–	–	0.920	0.889
	WKL	–	–	–	–	0.859
8	RL	0.721	0.673	0.633	0.628	0.843
	RI	–	0.665	0.743	0.724	0.901
	RV	–	–	0.557	0.555	0.765
	WE	–	–	–	0.950	0.912
	WKL	–	–	–	–	0.893

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	0.656	0.706	0.642	0.488	0.891
	RI	–	0.671	0.512	0.382	0.801
	RV	–	–	0.543	0.397	0.819
	WE	–	–	–	0.559	0.838
	WKL	–	–	–	–	0.644
4	RL	0.543	0.608	0.688	0.500	0.884
	RI	–	0.515	0.423	0.355	0.703
	RV	–	–	0.490	0.355	0.729
	WE	–	–	–	0.566	0.871
	WKL	–	–	–	–	0.662

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 Subclaims—Science

Grade	Subclaim	*Life Science	Earth and Space Science	Total Test
5	Physical Science	0.588	0.706	0.864
	Life Science*	–	0.674	0.830
	Earth and Space Science	–	–	0.932
8	Physical Science	0.756	0.761	0.922
	Life Science	–	0.735	0.915
	Earth and Space Science	–	–	0.902
11	Physical Science	0.718	0.715	0.895
	Life Science	–	0.730	0.908
	Earth and Space Science	–	–	0.903

*For Grade 5, the subclaim is Physical Science/Life Science.
Note. Correlations are provided only for the subclaims reported in Spring 2022.

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 the calibration of the Spring 2022 assessments supported standard setting for CMAS Science.

Calibration, equating, and scaling analyses for the operational and field test items were as follows for the Spring 2022 administration. The entire process was completed for each CMAS and CSLA assessment. All steps were independently replicated by at least two members of the Pearson psychometrics team to ensure accuracy.

- CMAS Mathematics
    - Operational items
        - All items had parameter estimates already equated to the base scale.
        - Used ISE to estimate student abilities
    - Embedded field test items
        - Used IRTPRO control files and IDM to obtain item parameter estimates of the operational and field test items
        - Used STUIRT to scale the field test items to the operational scale using the online operational items as the anchor set
        - Calculated item fit statistics and plotted expected vs. observed IRFs for each field test item

- CMAS ELA
    - Operational items
        - Used IRTPRO (SSI, Inc., 2011) control files and incomplete data matrix (IDM) to obtain the online operational item parameter estimates
        - Evaluated the consistency of scoring and stability of the anchor items
        - Used STUIRT (Kim & Kolen, 2004) to scale the 2022 operational items to the operational scale
        - Calculated item fit statistics and plotted expected vs. observed item response functions (IRFs) for each operational item
        - Used ISE (Chien & Shin, 2012) to estimate student abilities
    - Embedded field test items
        - Used IRTPRO control files and IDM to obtain item parameter estimates of the operational and field test items
        - Used STUIRT to scale the field test items to the operational scale using the online operational items as the anchor set
        - Calculated item fit statistics and plotted expected vs. observed IRFs for each field test item

- CSLA
    - Operational items
        - Used Winsteps (Linacre, 2011) control files and IDM to obtain the non-anchor operational item parameter estimates
        - Evaluated the stability of the anchor items to obtain the final anchor set
        - Used the final anchor set in Winsteps to scale the 2022 non-anchor items to the operational scale
        - Obtained item difficulty values, step deviation values, and item fit values for all items
        - Used Winsteps to estimate student abilities
    - Embedded field test items
        - Used Winsteps control files and IDM to scale the field test item parameter estimates to the operational scale by fixing the item parameter estimates of the operational items
        - Obtained field test item difficulty values, step deviation values, and item fit values for each field test item

- CMAS Science
    - Used IRTPRO control files and IDM to obtain the online operational and field test item parameter estimates

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*:

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

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:

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

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*:

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

where all parameters are as they were before, and d_{jv} is the category parameter for category *v* of item *j* and M_i 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.

The 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). The 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 proficiency level, θ_n (also referred to as "ability"), and the step difficulties, δ_{ij}, of the *m* steps in question *i* as follows:

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

10.2. Item Response Curves (IRCs) and Item Characteristic Curves (ICCs)

The IRFs of the 2PL, 3PL, and GPC IRT models 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 for dichotomous items are equal, but the IRC and ICF are different for polytomous items.

For example, consider Figure 10.1 that depicts a 2PL item that falls at approximately 0.85 on the ability (horizontal) 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. Another way of expressing this is that in a group of 100 students, all of whom 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. Sample 2PL Item Characteristic Curve (ICC)

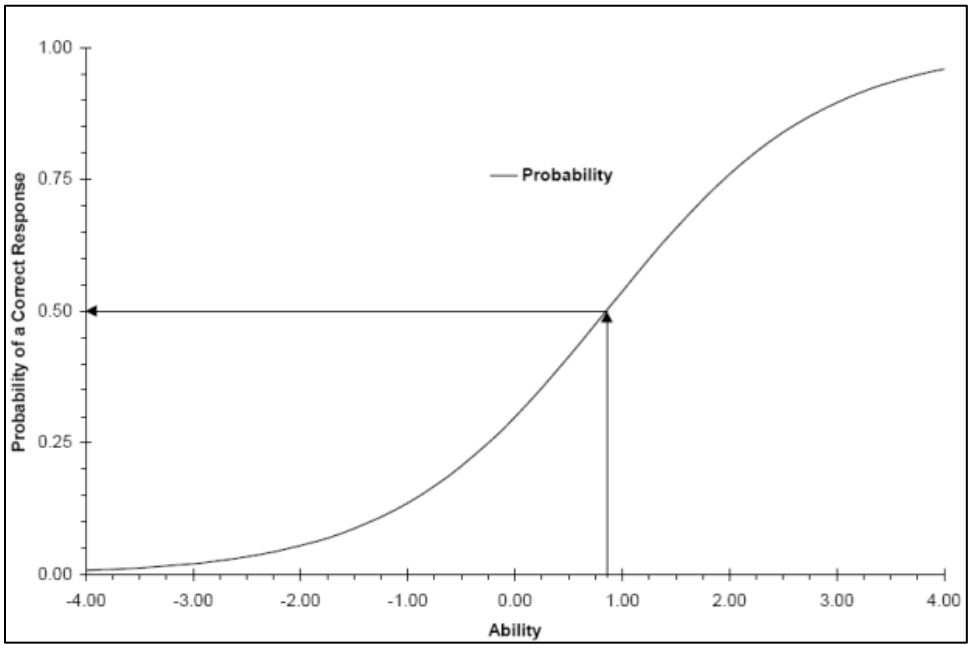


Figure 10.2 shows IRCs of obtaining a wrong answer or a right answer. The dotted-line curve (j=0) shows the probability of getting a score of 0, while the solid-line curve (j=1) shows the probability of getting a score of 1. The point at which the two curves cross indicates the transition point on the 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. Sample 2PL Item Response Curves (IRCs)

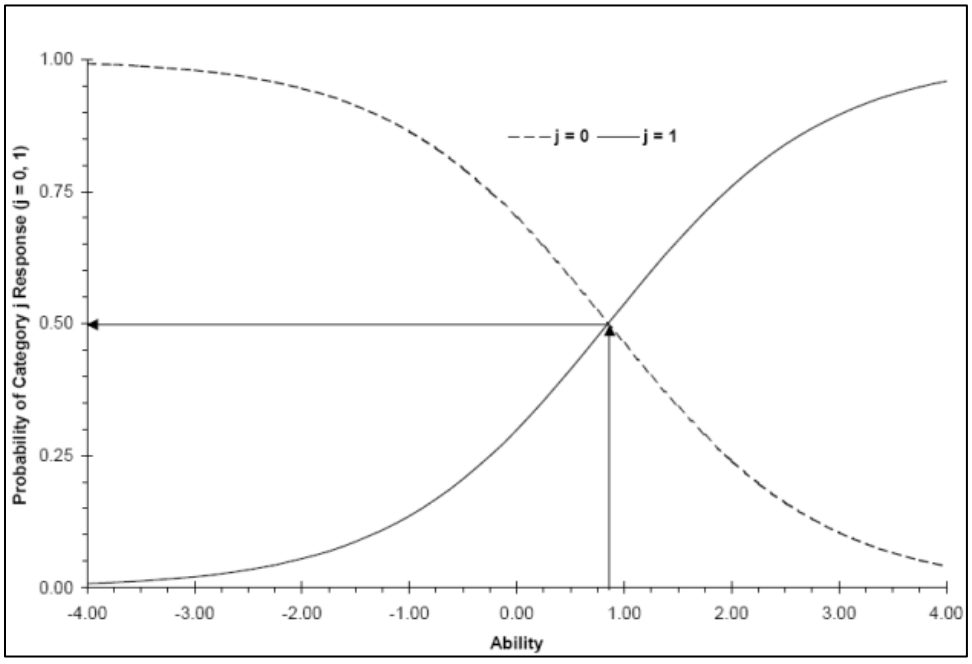
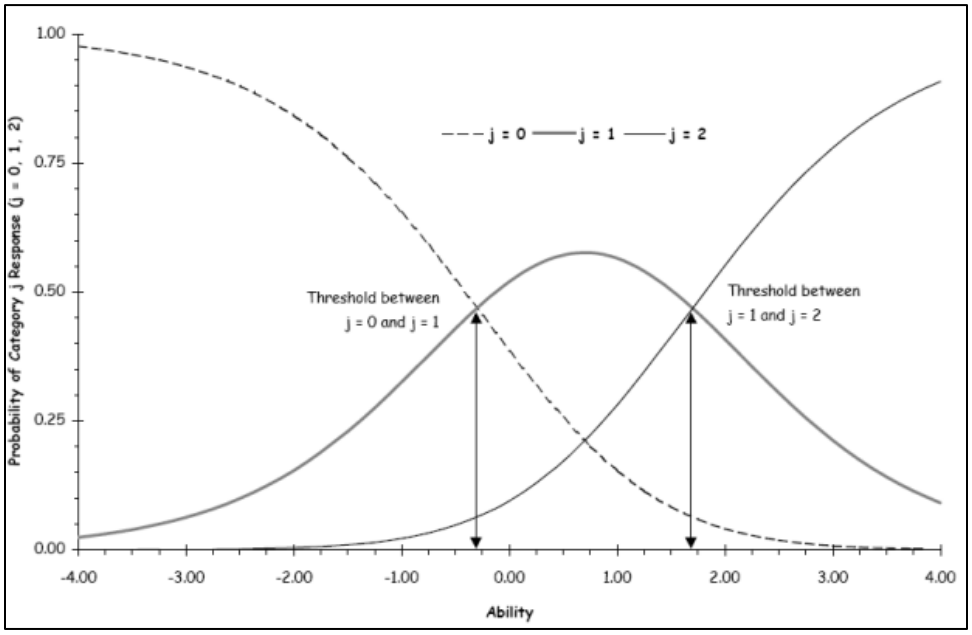


Figure 10.3 shows IRCs of obtaining each score category for a polytomously scored item. The dotted-line curve (j=0) shows the probability of getting a score of 0. Those of very low ability

(e.g., below -2) are likely to be in this category. Those receiving a 1 (partial credit) tend to fall in the middle range of abilities (the thick, solid-line curve, *j*=1). The final, thin, solid-line curve (*j*=2) represents the probability for those receiving scores of 2 (completely correct). Very high-ability students are more likely to be in this category, but there are still some of average and low ability who can get full credit for the item.

The points at which the lines cross have a similar interpretation as that 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. Item Response Curves (IRCs) for a 2-point Polytomous Item



10.3. Data Preparation

Prior to any analyses, several steps were completed as preparation:

- Verify the data file containing student responses and apply the exclusion rules.
- Complete a traditional item analysis (TRIAN) and adjudication, where applicable, on all items.
- Create incomplete data matrices (IDMs).

A TRIAN of all SR items was conducted prior to calibration. The purpose of this review is to use classical statistics to identify potential test administration and score issues. Items with one or more of the following characteristics are flagged:

- 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

A list of flagged items is communicated to the assessment specialists for review and confirmation that the correct key has been applied. Figure 10.4 presents a sample TRIAN report.

Figure 10.4. Sample Key Check (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 are put through an adjudication process. For each item, the frequency distribution of responses that are scored correctly is created, along with the frequency distribution of responses that are scored as incorrect. Assessment specialists review each response in the frequency reports and indicate whether the response should be scored as correct. The assessment specialists' indications are 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. Sample 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			0	56	0								
Item1		A_A2	0	1	0								
Item1		B_B1	0	12	0								
Item1		B_B1:A_A2	0	464	1								
Item1		B_B1:B_B2	0	1038	2								
Item1		B_B1:C_C2	0	844	2								
Item1		B_B1:D_D2	0	405	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_C1:D_D2	0	1510	3								
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.

10.4.1. Unidimensionality (Factor Analysis)

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 K presents the scree plots for the Spring 2022 administration. For most of the assessments, one factor explained most of the variance, which supports the use of a unidimensional IRT model, although the ELA and 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

Item fit results are provided in Appendix M. (The results are not included if a test is pre-equated.) 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:

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

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 the value could be evaluated using the chi-square distribution:

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

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 in the IRT framework, which places items and students on a common scale. To obtain item parameter estimates for CMAS ELA, the GPC model was applied to the items. IRTPRO was used for all calibrations, and all operational item parameters were estimated in a single calibration (i.e., concurrent calibration) for each assessment. For CSLA, the RPCM was applied to all items to obtain item parameter estimates. All operational items within a grade were also calibrated concurrently. Winsteps was used for all CSLA calibrations.

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.

For each CMAS Science assessment, a concurrent calibration was conducted to obtain item parameters for all the newly developed items. The calibration of the assessment administration supported the Fall 2022 standard setting and 2023 test construction.

10.6. Equating

Equating is used to place new 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. 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.

To maintain comparability with prior administrations, CMAS Mathematics and ELA item parameter estimates were equated to the established base scale used in 2017, and the CSLA item parameter estimates were equated to the Spring 2016 CSLA base scale. No equating was conducted for science in Spring 2022.

10.6.1. Mathematics

10.6.1.1. Pre-Equating Design

The Spring 2022 CMAS Mathematics assessments were equated to the base scales using an item pre-equating design, meaning all items had already been administered, with item parameters already estimated and placed onto the base scale. Students were scored based on these previously banked item parameter estimates. All operational items on these forms had been previously calibrated and equated to the base scale.[12] The forms were subsequently scored using these existing item parameters rather than performing a new calibration and equating. To help ensure the stability of item parameter estimates across administration, items were positioned as closely as possible to their positions when they were calibrated. To ensure that the assumptions of pre-equating were met, a post-equating check was performed using anchor sets identified during test construction. The results of this check were compared with the pre-equated results during a post-equating check.

10.6.1.2. Post-Equating Check

Because pre-equating relies on stronger assumptions than post-equating, an additional post-equating analysis was conducted and compared with the pre-equated results for the CMAS mathematics assessments. Large discrepancies between the two could suggest that pre-equating assumptions have not been met. Conversely, similarity between pre- and post-equated item parameters suggests that the pre-equated item parameters are appropriate for students taking the current form. The post-equating check followed the same procedures as those of the other post-equated assessments, using an anchor set for each assessment that was identified during test construction and that met the operational anchor test specifications.

Results of the post-equating check suggested that pre- and post-equated item parameters were quite similar. Appendix L compares the pre-equated and post-equated test characteristic curves (TCCs) for each assessment. The results of the check show that post-equated scores would have been highly similar to the pre-equated results. The high degree of similarity across the entire scale score range for each grade suggests that pre-equating assumptions were met and that the pre-equated item parameters were appropriate for this administration.

10.6.2. ELA

The Spring 2022 CMAS ELA assessments were calibrated and post-equated to the base scale following the procedures described below. The ELA assessments have historically been post-equated. 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 30% of a full form.

[12] Please refer to previous versions of the CMAS technical reports for information on these calibration and equating procedures, located on the CDE website at https://www.cde.state.co.us/assessment/cmas_coalt_techreport.

10.6.2.1. 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.

10.6.2.2. 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:

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

   where i = item, x = old form, y = new form, k = theta quadrature point, and g = theoretically weighted posterior theta distribution.

5. Compute the mean and standard deviation of the D^2 values.
6. Flag the items with a D^2 more than 2 standard deviations above the mean.

10.6.2.3. 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 one to four items from the anchor set, depending on grade. The final anchor sets for ELA represented between 29% and 47% of the unweighted total test points. 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.

10.6.2.4. 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.

10.6.2.5. 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:

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

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

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

10.6.2.6. 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.

10.6.3. 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.

10.6.3.1. 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.

10.6.3.2. 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 32% and 41%, respectively, of the unweighted total test points.

10.6.3.3. Equating Method

To obtain equated Rasch parameter estimates for the Spring 2022 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 that of operational equating, except that 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, with the exception of 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

10.8.1. Mathematics and ELA

Student ability was estimated using IRT pattern scoring based on student responses and the operational item parameter estimates. 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.

10.8.2. CSLA

After the item parameter estimates were obtained for the CSLA operational items, student abilities were estimated for each grade-level assessment by conducting an anchored calibration of the operational items' item parameter estimates. 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 Mathematics, ELA, and CSLA, 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. For CMAS ELA and 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.

10.10. Item-Level IRT Statistics

Appendix M 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 tables reflects whether an item was flagged for misfit based on Q1 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 N 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 2022 CMAS mathematics and ELA TCCs matched those from 2018 in terms of shape and position. The 2022 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 2022 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 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 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 ELA at grades 4, 6, and 8 and mathematics at grades 3, 5, and 7 were not required and only given on an opt-in basis, student participation was much lower in those grades in Spring 2021 than in previous years. As a result, the Spring 2022 prior year's overall test score did not come from the same content area for all grades. The use of cross subject prior year scores was evaluated. Correlations between ELA and mathematics scale scores indicated the two subjects are highly correlated. As a result, ELA prior year scores were used for mathematics grades 4, 6, and 8, and mathematics prior year scores were used for ELA grades 5 and 7. For Grade 3, no prior test scores were available, so those samples were matched on the demographic variables only.

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:

$$d = \frac{M_{group1} - M_{group2}}{SD_{pooled}}$$

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 the results were presented to CDE who made the final decision on whether to make an adjustment for mode differences for each assessment. Table 10.1 presents the mode effect sizes for mathematics and ELA from the Spring 2022 administration.

Based on evaluation of the effect sizes, mode adjustments were made for mathematics Grades 7 and 8 and ELA Grades 5–8. 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 the paper-based mathematics assessments from the prior administration, mode adjustments from that prior administration were applied to those forms.

Table 10.1. Online vs. Paper Comparability Mode Effect Sizes

	Mathematics		ELA	
Grade	N	Effect Size	N	Effect Size
3	4,484	0.00	4,669	-0.06
4	3,316	0.02	3,415	-0.05
5	3,185	0.06	3,250	-0.10
6	3,000	-0.08	2,761	-0.27
7	2,873	-0.10	2,558	-0.21
8	2,691	-0.12	2,302	-0.25

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 H presents the coefficient alpha estimates by demographic subgroup. The internal consistency values for the overall test ranged from 0.83 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—Mathematics

Grade	Overall	Subclaim A	Subclaim B	Subclaim C	Subclaim D
3	0.92	0.86	0.79	0.67	0.68
4	0.91	0.88	0.61	0.73	0.51
5	0.93	0.88	0.72	0.74	0.66
6	0.90	0.84	0.62	0.74	0.62
7	0.90	0.82	0.55	0.77	0.56
8	0.90	0.84	0.43	0.76	0.67

Table 11.2. Coefficient Alpha—ELA

Grade	Overall	Reading: Literary Text	Reading: Informational Text	Reading: Vocabulary	Writing: Written Expression	Writing: Knowledge and Use of Language Conventions	Reading	Writing
3	0.90	0.74	0.74	0.66	0.67	0.76	0.88	0.75
4	0.90	0.74	0.76	0.67	0.70	0.77	0.88	0.81
5	0.87	0.70	0.69	0.45	0.68	0.73	0.84	0.79
6	0.89	0.76	0.72	0.52	0.74	0.78	0.87	0.83
7	0.90	0.78	0.72	0.61	0.81	0.81	0.88	0.86
8	0.90	0.75	0.75	0.58	0.88	0.87	0.88	0.88

Table 11.3. Coefficient Alpha—CSLA

Grade	Overall	Reading: Literary Text	Reading: Informational Text	Reading: Vocabulary	Writing: Written Expression	Writing: Knowledge and Use of Language Conventions	Reading	Writing
3	0.88	0.78	0.66	0.69	0.61	0.82	0.88	0.73
4	0.83	0.71	0.50	0.59	0.54	0.73	0.81	0.69

Table 11.4. Coefficient Alpha—Science

Grade	Overall	Physical Science	*Life Science	Earth and Space Science
5	0.86	0.69	0.58	0.74
8	0.89	0.72	0.79	0.64
11	0.89	0.69	0.71	0.76

*For Grade 5, the subclaim 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—Mathematics

Grade	Subclaim A	Subclaim B	Subclaim C	Subclaim D
3	1.97	1.16	1.53	1.30
4	2.11	1.10	1.67	1.83
5	2.07	1.22	1.35	1.47
6	1.88	1.48	1.51	1.44
7	1.94	1.24	1.34	1.43
8	2.04	1.40	1.14	1.31

Table 11.6. SEM—ELA

Grade	Reading: Literary Text	Reading: Informational Text	Reading: Vocabulary	Writing: Written Expression	Writing: Knowledge and Use of Language Conventions	Writing
3	1.83	1.83	1.56	2.57	0.59	2.68
4	1.99	2.11	1.48	3.20	0.69	3.08
5	2.25	1.84	1.51	3.46	0.69	3.32
6	2.21	2.30	1.54	3.52	0.73	3.42
7	1.90	2.36	1.67	3.26	0.73	3.38
8	2.09	2.41	1.78	3.04	0.72	3.69

Table 11.7. SEM—CSLA

Grade	Reading: Literary Text	Reading: Informational Text	Reading: Vocabulary	Writing: Written Expression	Writing: Knowledge and Use of Language Conventions	Writing
3	1.95	1.83	1.54	3.71	0.77	3.71
4	2.17	2.05	1.51	4.69	0.94	4.50

Table 11.8. SEM—Science

Grade	Physical Science	*Life Science	Earth and Space Science
5	1.80	1.50	2.01
8	1.60	1.79	1.89
11	1.25	1.96	1.60

*For Grade 5, the subclaim 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 N 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, ELA, and 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

Content Area	Grade	Accuracy	Consistency	PChance	Kappa
Mathematics	3	0.73	0.63	0.22	0.53
	4	0.75	0.65	0.24	0.54
	5	0.76	0.67	0.23	0.57
	6	0.73	0.63	0.24	0.51
	7	0.75	0.65	0.26	0.53
	8	0.72	0.63	0.23	0.52

Content Area	Grade	Accuracy	Consistency	PChance	Kappa
ELA	3	0.72	0.62	0.25	0.50
	4	0.72	0.61	0.25	0.48
	5	0.86	0.81	0.28	0.73
	6	0.71	0.61	0.25	0.48
	7	0.70	0.60	0.22	0.49
	8	0.87	0.81	0.23	0.75
CSLA	3	0.74	0.64	0.26	0.51
	4	0.72	0.62	0.31	0.44

Table 11.11. Accuracy of Cut Scores

Content Area	Grade	*Partially Met Expectations* Cut	*Approached Expectations* Cut	*Met Expectations* Cut	*Exceeded Expectations* Cut
Mathematics	3	0.94	0.92	0.91	0.89
	4	0.94	0.91	0.92	0.97
	5	0.95	0.92	0.92	0.92
	6	0.93	0.90	0.92	0.96
	7	0.95	0.90	0.92	0.97
	8	0.92	0.90	0.92	0.94
ELA	3	0.93	0.91	0.90	0.94
	4	0.95	0.92	0.90	0.91
	5	0.98	0.95	0.95	0.94
	6	0.95	0.91	0.90	0.91
	7	0.94	0.91	0.90	0.85
	8	0.97	0.91	0.96	0.88
CSLA	3	0.89	0.80	0.83	0.98
	4	0.92	0.79	0.86	0.99

Table 11.12. Consistency of Cut Scores

Content Area	Grade	*Partially Met Expectations* Cut	*Approached Expectations* Cut	*Met Expectations* Cut	*Exceeded Expectations* Cut
Mathematics	3	0.92	0.89	0.88	0.88
	4	0.91	0.87	0.88	0.96
	5	0.92	0.89	0.89	0.91
	6	0.90	0.86	0.88	0.95
	7	0.92	0.86	0.89	0.96
	8	0.89	0.87	0.88	0.93
ELA	3	0.90	0.87	0.86	0.93
	4	0.93	0.89	0.86	0.88
	5	0.97	0.94	0.93	0.94
	6	0.93	0.88	0.86	0.89
	7	0.92	0.87	0.87	0.83
	8	0.96	0.94	0.94	0.88
CSLA	3	0.84	0.72	0.77	0.96
	4	0.89	0.70	0.80	0.98

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. Appendix O 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:

- 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 (Kappa), $KAPPA = \frac{E([X_1 - Y_1]^2)}{E([X_1 - Y_2]^2)}$, 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)
- Standardized mean differences (MD): $\bar{Z} = \frac{|\bar{X}_{R_1} - \bar{X}_{R_2}|}{\sqrt{\frac{sd^2_{R_1} + sd^2_{R_2}}{2}}}$
- Correlations (CORR): $\bar{Z} = \frac{|\bar{X}_{R_1} - \bar{X}_{R_2}|}{\sqrt{\frac{sd^2_{R_1} + sd^2_{R_2}}{2}}}$

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 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 ESs or EOs. Test items are reviewed for adherence to universal design principles to maximize the participation of the widest possible range of students.

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 (including CSLA) item and test development:

- The item development process ensures that the mathematics and ELA (including 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 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 achievement 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: Item Development. 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: Test Construction provides details on the construction of the operational assessment forms, which demonstrates that all test forms for mathematics and ELA adhered to the same test design used in previous years or were previously used operationally. Science was a new assessment administered in Spring 2022; as such, validity evidence based on test content for science will continue to be documented in future iterations of this technical report as the test continues to be developed.

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 test blueprints.

The CMAS assessments provide a full summative test score and a Reading score, as well as percent of points earned scores for Writing and mathematics, ELA, and science 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 subclaim 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 1.4. 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 and ELA 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 mathematics and ELA scale scores were fairly high; these values are also very close to the 2018 values.

Table 12.1. Correlations Between Mathematics and ELA Scale Scores

Grade	N	Correlation
3	54862	0.79
4	55478	0.77
5	57029	0.74
6	55548	0.76
7	54818	0.76
8	52199	0.75

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 with Colorado students in May 2013. Students attempted a variety of item types on the TestNav platform and were asked to "think-aloud" as they worked through each item. Students showed a high degree of facility in responding to the items, and only a small bit of supplemental training was speculated to be needed to acquaint them with the tools and navigation of the TestNav interface. Surveys were given to the students after completion of the assessment, which included a question that asked them to indicate whether they preferred paper or computer-based tests. Most students indicated that they preferred the computer-based version, and many commented that it had been an enjoyable experience. For a full report on the cognitive labs, see the *2013–2014 CMAS Technical 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 "with the expectation that some benefit will be realized from the intended use of the scores" (AERA et al., 2014), 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.

Table 12.2 presents the percentage of students who have reached proficiency on the CMAS assessments over the years. The CMAS Mathematics and ELA assessments have been administered to Colorado students since Spring 2015, and CSLA has been administered since Spring 2016. While CMAS Science has been administered since Spring 2014, it is not included in the table because a new CMAS science assessment based on new standards was administered for the first time in Spring 2022.

As shown in the table, student performance has improved since the first administration with the exception of Grade 6 mathematics and Grades 3 and 4 CSLA. The decrease in 2021 of the percent of students meeting or exceeding in the required grades was expected given the learning disruptions caused by COVID-19 in 2020 and 2021. The pandemic's continued impact on student learning opportunities should be considered when interpreting the 2022 performance results. There have also been changes in the available assessments by grade for Grades 7 and 8 mathematics across administrations, so comparisons across years for those grades are not included.

Table 12.2. Student Performance Over Time

Content Area	Grade	First Administration *%Met or Exceeded*	2019 *%Met or Exceeded*	2021 *%Met or Exceeded*	2022 *%Met or Exceeded*	% Change, First Administration to 2019	% Change, First Administration to 2021	% Change, First Administration to 2022
Mathematics	3	36.7	41.0	–	39.4	4.3	–	2.7
	4	30.2	33.6	28.5	30.7	3.4	-1.7	0.5
	5	30.1	35.7	–	34.9	5.6	–	4.8
	6	31.7	29.5	24.1	26.3	-2.2	-7.6	-5.4

Content Area	Grade	First Administration *%Met or Exceeded*	2019 *%Met or Exceeded*	2021 *%Met or Exceeded*	2022 *%Met or Exceeded*	% Change, First Administration to 2019	% Change, First Administration to 2021	% Change, First Administration to 2022
ELA	3	38.2	41.3	39.1	40.7	3.1	0.9	2.5
	4	41.7	48.0	–	44.1	6.3	–	2.4
	5	40.5	48.4	47.2	45.4	7.9	6.7	4.9
	6	39.1	43.6	–	43.0	4.5	–	3.9
	7	41.0	46.5	42.6	41.8	5.5	1.6	0.8
	8	40.9	46.9	–	43.9	5.9	–	3.0
CSLA	3	22.0	27.5	15.4	19.8	5.5	-6.6	-2.2
	4	13.9	19.1	–	13.7	5.2	–	-0.2

Note. The first administration for mathematics ELA was Spring 2015, and the first administration for CSLA was Spring 2016. Performance results are not included for the Spring 2021 mathematics, ELA, and CSLA opt-in grades.

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: Item Development)
- Items are reviewed by educators for potential issues which could impact the performance of student groups prior to field testing (Chapter 3: Item Development).
- 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*. https://parcc-assessment.org/wp-content/uploads/2017/12/PARCC_PLS_TechReport_011316_toPARCC-final.pdf

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.

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.

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

Scientific Software International (SSI), Inc. (2011). *IRTPRO [computer program].*

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.

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	Mini Cluster	Standalone
Physical	35	9	6–12	4–10
Physical/Life	24	9	3–6	2–5
Earth and Space	41	9	6–12	9–15
Science and Engineering Practices	65–75	–	–	–
SR and TE	53	–	–	–
CR	47	–	–	–
Total	100	27	24	21

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

Standard/Item Type	%Total Test Score Points	#Points		
		Cluster	Mini Cluster	Standalone
Physical	35	9	9–15	3–9
Physical/Life	34	9	9–15	4–10
Earth and Space	29	9	6–9	4–7
Science and Engineering Practices	65–74	–	–	–
SR and TE	53	–	–	–
CR	47	–	–	–
Total	100	27	30	20

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

Standard/Item Type	%Total Test Score Points	#Points		
		Cluster	Mini Cluster	Standalone
Physical	46	9	12	7
Physical/Life	32	9	6–12	4–10
Earth and Space	31	9	6–12	3–9
Science and Engineering Practices	65–74	–	–	–
SR and TE	53	–	–	–
CR	47	–	–	–
Total	100	27	30	17

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: ELA and CSLA Scoring Rubrics

Grade 3

CMAS Scoring Rubric for Prose Constructed Response Items

COLORADO
Department of Education

Research Simulation Task (RST) and Literary Analysis Task (LAT)

Construct Measured	Score Point 3	Score Point 2	Score Point 1	Score Point 0
Reading Comprehension and Written Expression	The student response • demonstrates **full comprehension** by providing an **accurate** explanation/description/comparison; • addresses the prompt and provides **effective** development of the topic that is **consistently appropriate** to task, purpose, and audience; • uses **clear** reasoning supported by **relevant**, text- based evidence in the development of the topic; • is **effectively** organized with **clear and coherent** writing; • uses language **effectively** to clarify ideas.	The student response • demonstrates **comprehension** by providing a **mostly accurate** explanation/description/comparison; • addresses the prompt and provides **some** development of the topic that is **generally appropriate** to task, purpose, and audience; • uses reasoning and **relevant**, text-based evidence in the development of the topic; • is organized with **mostly clear and coherent** writing; • uses language in a way that is **mostly effective** to darify ideas.	The student response • demonstrates **limited comprehension**; • addresses the prompt and provides **minimal** development of the topic that is **limited in its appropriateness** to task, purpose, and audience; • uses **limited** reasoning and text-based evidence; • demonstrates **limited** organization and coherence; • uses language to express ideas with **limited** clarity.	The student response • **does not** demonstrate comprehension; • is **undeveloped and/or inappropriate** to the task, purpose, and audience; • includes **little to no** text-based evidence; • **lacks** organization and coherence; • **does not** use language to express ideas with clarity.
Knowledge of Language and Conventions	The student response to the prompt demonstrates **full command** of the conventions of standard English at an appropriate level of complexity. There **may be a few minor errors** in mechanics, grammar, and usage, but **meaning is clear**.	The student response to the prompt demonstrates **some command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **occasionally impede understanding**, but the **meaning is generally clear**.	The student response to the prompt demonstrates **limited command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **often impede understanding.**	The student response to the prompt **does not demonstrate command** of the conventions of standard English at the appropriate level of complexity. **Frequent and varied errors** in mechanics, grammar, and usage **impede understanding**.

Grade 3 CMAS Scoring Rubric for Prose Constructed Response Items

Narrative Task (NT)

Construct Measured	Score Point 3	Score Point 2	Score Point 1	Score Point 0
Written Expression	The student response • is **effectively** developed with narrative elements and is **consistently appropriate** to the task; • is **effectively** organized with **clear and coherent** writing • uses language **effectively** to clarify ideas.	The student response • is developed with **some** narrative elements and is **generally appropriate** to the task; • is organized with **mostly coherent** writing; • uses language in a way that is **mostly effective** to clarify ideas.	The student response • is **minimally** developed with **few** narrative elements and is **limited in its appropriateness** to the task; • demonstrates **limited** organization and coherence; • uses language to express ideas with **limited** darity.	The student response • is **undeveloped** and/or **inappropriate** to the task; • **lacks** organization and coherence; • **does not** use language to express ideas with clarity.
Knowledge of Language and Conventions	The student response to the prompt demonstrates **full command** of the conventions of standard English at an appropriate level of complexity. There may be a **few minor errors** in mechanics, grammar, and usage, but **meaning is clear**.	The student response to the prompt demonstrates **some command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **occasionally impede understanding**, but the **meaning is generally clear**.	The student response to the prompt demonstrates **limited command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **often impede understanding**.	The student response to the prompt **does not demonstrate command** of the conventions of standard English at the appropriate level of complexity. **Frequent and varied errors** in mechanics, grammar, and usage **impede understanding**.

NOTE:

- The reading dimension is not scored for elicited narrative stories.
- Per the CCSS, narrative elements in grades 3-5 may include: establishing a situation; organizing a logical event sequence; describing scenes, objects, or people; developing characters' personalities; and using dialogue as appropriate.
- The elements of organization to be assessed are expressed in the grade-level standards W1-W3.

Developed collaboratively with PARCC.

Grades 4 and 5

CMAS Scoring Rubric for Prose Constructed Response Items

COLORADO
Department of Education

Research Simulation Task (RST) and Literary Analysis Task (LAT)

Construct Measured	Score Point 4	Score Point 3	Score Point 2	Score Point 1	Score Point 0
Reading Comprehension and Written Expression	The student response • demonstrates **full comprehension** of ideas stated explicitly and/or inferentially by providing an **accurate** analysis; • addresses the prompt and provides **effective** development of the topic that is **consistently appropriate** to task, purpose, and audience; • uses **clear** reasoning supported by **relevant**, text-based evidence in the development of the topic; • is **effectively** organized with **clear and coherent** writing; • uses language **effectively** to clarify ideas.	The student response • demonstrates **comprehension** of ideas stated explicitly and/or inferentially by providing a **mostly accurate** analysis; • addresses the prompt and provides **mostly effective** development of the topic that is **appropriate** to task, purpose, and audience; • uses **mostly clear** reasoning supported by **relevant** text-based evidence in the development of the topic; • is organized with **mostly clear and coherent** writing • uses language that is **mostly effective** to clarify ideas.	The student response • demonstrates **basic comprehension** of ideas stated explicitly and/or inferentially by providing a **generally accurate** analysis; • addresses the prompt and provides **some** development of the topic that is **somewhat appropriate** to task, purpose, and audience; • uses **some** reasoning and text-based evidence in the development of the topic; • demonstrates **some** organization with **somewhat** coherent writing; • uses language to express ideas with **some** clarity.	The student response • demonstrates **limited comprehension** of ideas by providing a **minimally accurate** analysis; • addresses the prompt and provides **minimal** development of the topic that is **limited in its appropriateness** to task, purpose, and audience; • uses **limited** reasoning and text-based evidence; • demonstrates **limited** organization and coherence; • uses language to express ideas with **limited** clarity.	The student response • demonstrates **no comprehension** of ideas by providing an **inaccurate or no** analysis. • is **undeveloped and/or inappropriate** to the task, purpose, and audience; • includes **little to no** text- based evidence; • **lacks** organization and coherence; • **does not** use language to express ideas with clarity.
Knowledge of Language and Conventions		The student response to the prompt demonstrates **full command** of the conventions of standard English at an appropriate level of complexity. There **may be a few minor errors** in mechanics, grammar, and usage, but **meaning is clear**.	The student response to the prompt demonstrates **some command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **occasionally impede understanding**, but the **meaning is generally clear**.	The student response to the prompt demonstrates **limited command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **often impede understanding.**	The student response to the prompt **does not demonstrate command** of the conventions of standard English at the appropriate level of complexity. **Frequent and varied errors** in mechanics, grammar, and usage **impede understanding**.

Grades 4 and 5 CMAS Scoring Rubric for Prose Constructed Response Items

Narrative Task (NT)

Construct Measured	Score Point 3	Score Point 2	Score Point 1	Score Point 0
Written Expression	The student response • is **effectively** developed with narrative elements and is **consistently appropriate** to the task; • is **effectively** organized with **clear and coherent** writing • uses language **effectively** to clarify ideas.	The student response • is developed with **some** narrative elements and is **generally appropriate** to the task; • is organized with **mostly coherent** writing; • uses language in a way that is **mostly effective** to clarify ideas.	The student response • is **minimally** developed with **few** narrative elements and is **limited in its appropriateness** to the task; • demonstrates **limited** organization and coherence; • uses language to express ideas with **limited** darity.	The student response • is **undeveloped** and/or **inappropriate** to the task; • **lacks** organization and coherence; • **does not** use language to express ideas with clarity.
Knowledge of Language and Conventions	The student response to the prompt demonstrates **full command** of the conventions of standard English at an appropriate level of complexity. There may be a **few minor errors** in mechanics, grammar, and usage, but **meaning is clear**.	The student response to the prompt demonstrates **some command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **occasionally impede understanding**, but the **meaning is generally clear**.	The student response to the prompt demonstrates **limited command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **often impede understanding**.	The student response to the prompt **does not demonstrate command** of the conventions of standard English at the appropriate level of complexity. **Frequent and varied errors** in mechanics, grammar, and usage **impede understanding**.

NOTE:

- The reading dimension is not scored for elicited narrative stories.
- Per the CCSS, narrative elements in grades 3-5 may include: establishing a situation; organizing a logical event sequence; describing scenes, objects, or people; developing characters' personalities; and using dialogue as appropriate.
- The elements of organization to be assessed are expressed in the grade-level standards W1-W3.

Developed collaboratively with PARCC.

Grades 6 through 8

CMAS Scoring Rubric for Prose Constructed Response Items

COLORADO
Department of Education

Research Simulation Task (RST) and Literary Analysis Task (LAT)

Construct Measured	Score Point 4	Score Point 3	Score Point 2	Score Point 1	Score Point 0
Reading Comprehension and Written Expression	The student response • demonstrates **full** comprehension of ideas stated explicitly and inferentially by providing an **accurate** analysis; • addresses the prompt and provides **effective and comprehensive** development of the claim or topic that is **consistently appropriate** to task, purpose, and audience; • uses **clear** reasoning supported by **relevant** text-based evidence in the development of the claim or topic; • is **effectively** organized with **clear and coherent** writing; • establishes and maintains an **effective** style.	The student response • demonstrates comprehension of ideas stated explicitly and/or inferentially by providing a **mostly accurate** analysis; • addresses the prompt and provides **mostly effective** development of claim or topic that is **mostly appropriate** to task, purpose, and audience; • uses **mostly clear** reasoning supported by **relevant** text-based evidence in the development of the claim or topic; • is organized with **mostly clear and coherent** writing; • establishes and maintains a **mostly effective** style.	The student response • demonstrates **basic** comprehension of ideas stated explicitly and/or inferentially by providing a **generally accurate** analysis; • addresses the prompt and provides **some** development of claim or topic that is **somewhat appropriate** to task, purpose, and audience; • uses **some** reasoning and text-based evidence in the development of the claim or topic; • demonstrates **some** organization with **somewhat** coherent writing; • has a style that is **somewhat** effective.	The student response • demonstrates **limited** comprehension of ideas stated explicitly and/or inferentially by providing a **minimally accurate** analysis; • addresses the prompt and provides **minimal** development of claim or topic that is **limited in its appropriateness** to task, purpose, and audience; • uses **limited** reasoning and text-based evidence; • demonstrates **limited** organization and coherence; • has a style that is **minimally** effective.	The student response • demonstrates **no** comprehension of ideas by providing an **inaccurate or no** analysis; • is **undeveloped and/or inappropriate** to task, purpose, and audience; • includes **little to no** text-based evidence; • **lacks** organization and coherence; • has an **inappropriate** style.
Knowledge of Language and Conventions		The student response to the prompt demonstrates **full command** of the conventions of standard English at an appropriate level of complexity. There **may be a few minor errors** in mechanics, grammar, and usage, but **meaning is clear**.	The student response to the prompt demonstrates **some command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **occasionally impede understanding**, but the **meaning is generally clear**.	The student response to the prompt demonstrates **limited command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **often impede understanding.**	The student response to the prompt **does not demonstrate command** of the conventions of standard English at the appropriate level of complexity. **Frequent and varied errors** in mechanics, grammar, and usage **impede understanding**.

Grades 6 through 8 CMAS Scoring Rubric for Prose Constructed Response Items

Narrative Task (NT)

Construct Measured	Score Point 4	Score Point 3	Score Point 2	Score Point 1	Score Point 0
Written Expression	The student response • is **effectively** developed with narrative elements and is **consistently appropriate** to the task; • is **effectively** organized with **clear and coherent** writing • establishes and maintains an **effective** style.	The student response • is **mostly effectively** developed with narrative elements and is **mostly appropriate** to the task; • is organized with **mostly clear and coherent** writing • establishes and maintains a **mostly effective** style.	The student response • is developed with **some** narrative elements and is **generally appropriate** to the task; • demonstrates **some** organization with **somewhat coherent** writing; • has a style that is **somewhat** effective.	The student response • is **minimally** developed with **few** narrative elements and is **limited in its appropriateness** to the task; • demonstrates **limited** organization and coherence; • has a style that has **limited** effectiveness.	The student response • is **undeveloped and/or inappropriate** to the task; • **lacks** organization and coherence; • has an **inappropriate** style.
Knowledge of Language and Conventions		The student response to the prompt demonstrates **full command** of the conventions of standard English at an appropriate level of complexity. There may be a **few minor errors** in mechanics, grammar, and usage, but **meaning is clear**.	The student response to the prompt demonstrates **some command** of the conventions of standard English at an appropriate level of complexity. There **may** be errors in mechanics, grammar, and usage that **occasionally impede understanding**, but the **meaning is generally clear**.	The student response to the prompt demonstrates **limited command** of the conventions of standard English at an appropriate level of complexity. There **may** be errors in mechanics, grammar, and usage that **often impede understanding**.	The student response to the prompt **does not demonstrate command** of the conventions of standard English at the appropriate level of complexity. **Frequent and varied errors** in mechanics, grammar, and usage **impede understanding**.

NOTE:

- The reading dimension is not scored for elicited narrative stories.
- The elements of coherence, clarity, and cohesion to be assessed are expressed in the grade-level standards 1-4 for writing.
- Tone is not assessed in grade 6.
- Per the CCSS, narrative elements in grades 3-5 may include: establishing a situation; organizing a logical event sequence; describing scenes, objects, or people; developing characters' personalities; and using dialogue as appropriate. In grades 6-8, narrative elements may include, in addition to the grades 3-5 elements: establishing a context, situating events in a time and place, developing a point of view, and developing characters' motives. The elements to be assessed are expressed in grade-level standards 3 for writing.

Developed collaboratively with PARCC.

Appendix D: Sample Student Performance Reports

-*- Demonstration Powered by OpenText Exstream 02/02/2022, Version 16.6.31.64 bit -*-

Confidential Student Performance Report

Colorado Measures of Academic Success

Student: **FIRSTNAME038 D. (CHUCK) LASTNAME038**

SASID: **2021040119** Birthdate: **01/07/2007**
School: **SAMPLE SCHOOL NAME (4444)**
District: **SAMPLE DISTRICT NAME (5555)**

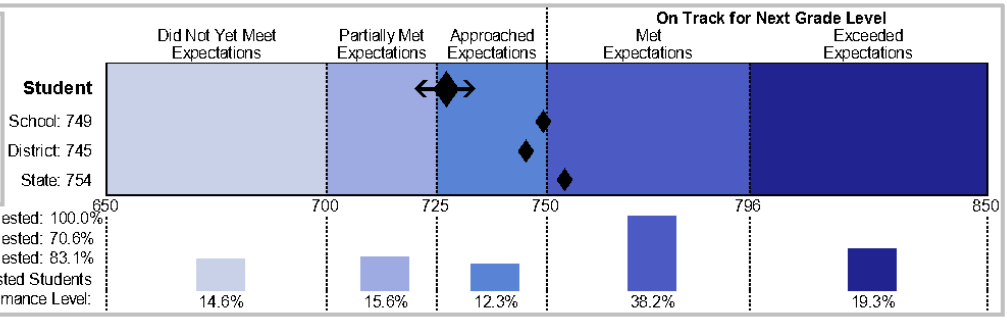
Spring 2022

Mathematics — Grade 4

This CMAS report provides information about your student's mastery of skills and concepts in the academic standards which are the basis for instruction in Colorado schools. Your student's performance on this test is represented by a scale score, a performance level, and a percentile rank.

- Scale scores are represented by diamonds on the graph. The arrows around your student's diamond show the range of scores your student would likely receive if the assessment were taken multiple times. Scale scores can be compared across years.
- School, district, and state information allows you to compare your student's performance to the performance of others. The percentage of students in each performance level across the state is reported below the graph.
- Peformance levels are separated by dotted lines.
- You are encouraged to discuss this report with your student's teacher.

Your Student's Score
727
Approached Expectations
32nd Percentile



Percent of School Students Tested: 100.0%
Percent of District Students Tested: 70.6%
Percent of CO Students Tested: 83.1%
Percent of CO Tested Students by Performance Level:

Performance Level Descriptor* - Approached Expectations

CHUCK **Approached Expectations** and may benefit from additional support to meet expectations at the next grade level. Students in this level typically demonstrate the following:

Major, Additional & Supporting Content

- Solve scaffolded problems involving comparison using multiplication.
- Solve two-step word problems with at least one two- or three-digit number. Generate a pattern from a given rule. With scaffolding, read, write and compare three-digit whole numbers and round to any place. Determine whether a whole number in the range of 1-100 is prime or composite with scaffolding.
- Recognize that decimals and fractions must refer to the same whole in order to compare.
- Given a model, compare fractions using benchmarks. Solve simple fraction comparison word problems. Use decimal notations for fractions. Multiply a fraction by a whole number using models, decompose a fraction into a sum of fractions with like denominators, and record using an equation.
- Convert units from larger to smaller units within the same system. Make a line plot to display data of measurements with like denominators of 2 or 4. Use a protractor to measure angles. Use criteria to classify quadrilaterals and triangles.
- Recognize that a whole number is a multiple of each of its factors, and find factor pairs or determine multiples of whole numbers.

Expressing Mathematical Reasoning

- Communicate reasoning that may include minor calculation errors. Provide a numerically complete response with partial justification, and evaluate the validity of claims made by others.

Modeling & Application

- Draw conclusions by illustrating the relationship between important quantities, modifying a model, or interpreting mathematical results in a simplified context.

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 the full version of the PLDs, visit https://coassessments.com/parentsandguardians.

*Adapted from ilClassroom in Action's Performance Level Summaries

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

Page 1 of 2

02022022-Z9999999-5555-4444 - 0000000

-*- Demonstration Powered by OpenText Exstream 02/02/2022, Version 16.6.31 64-bit -*-

FIRSTNAME038 D. (CHUCK) LASTNAME038

Mathematics

Confidential

Subclaim Performance

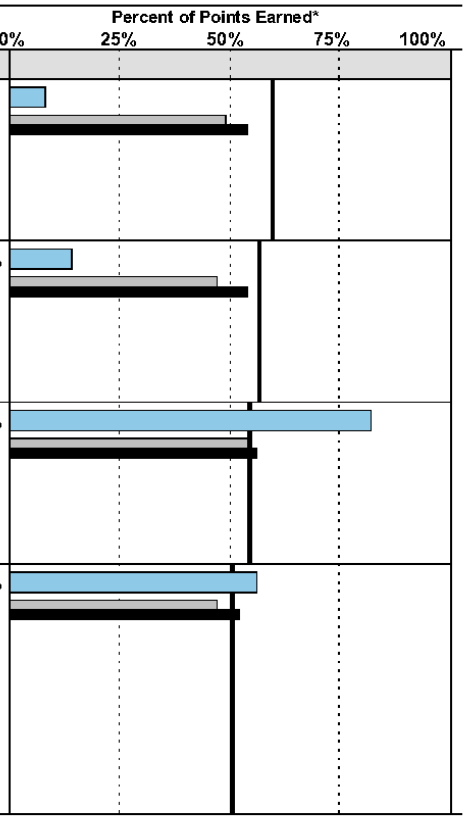
The top bar in each of the other graphs shows the percent of points your student earned for each of the four mathematics assessment subclaims.

District Averages are provided for comparison.

State Averages are provided for comparison.

The dark vertical line indicates the average percent of points earned by students who just crossed into the Met Expectations performance level on the overall Mathematics test.

	Points Possible	Percent of Points Earned*
Mathematics - Refer to page 1 for participation rates.		
Major Content Students solve problems involving addition, subtraction, multiplication and division, place value, fraction comparisons, and addition and subtraction of fractions with same denominators.	24	8%
Additional & Supporting Content Students solve problems involving number and shape patterns, simple measurement conversions, angle measurements, geometric shapes classification, and representations of data.	7	14%
Expressing Mathematical Reasoning Students create and justify logical mathematical solutions and analyze and correct the reasoning of others.	11	82%
Modeling & Application Students solve real-world problems, represent and solve problems with symbols, reason quantitatively, and strategically use appropriate tools.	9	56%



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

For more information about the standards included in this assessment, please visit the Colorado Department of Education's website at http://www.cde.state.co.us/comath/statestandards.

Page 2 of 2

Confidential Student Performance Report

Colorado Measures of Academic Success

Student: **FIRSTNAME008 (ROBERTA) LASTNAME008**
SASID: **2021040238** Birthdate: **01/14/2007**
School: **SAMPLE SCHOOL NAME (4444)**
District: **SAMPLE DISTRICT NAME (5555)**

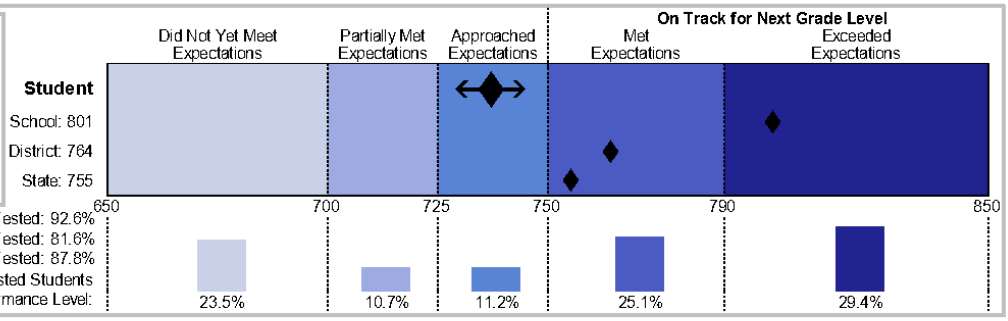
Spring 2022

English Language Arts/Literacy — Grade 4

This CMAS report provides information about your student's mastery of skills and concepts in the academic standards which are the basis for instruction in Colorado schools. Your student's performance on this test is represented by a scale score, a performance level, and a percentile rank.

- Scale scores are represented by diamonds on the graph. The arrows around your student's diamond show the range of scores your student would likely receive if the assessment were taken multiple times. Scale scores can be compared across years.
- School, district, and state information allows you to compare your student's performance to the performance of others. The percentage of students in each performance level across the state is reported below the graph.
- Performance levels are separated by dotted lines.
- You are encouraged to discuss this report with your student's teacher.

Your Student's Score
737
Approached Expectations
41st Percentile



Performance Level Descriptor - Approached Expectations

ROBERTA **Approached Expectations** and may benefit from additional support to meet expectations at the next grade level. Students in this level typically demonstrate the following:

Reading

- With very complex text: the ability to ask and/or answer questions with minimal accuracy, showing minimal understanding of the text when referring to explicit details and examples in the text.
- With moderately complex text: the ability to be generally accurate when asking and/or answering questions, showing basic understanding of the text when referring to explicit details and examples in the text.
- With readily accessible text: the ability to be mostly accurate when asking and/or answering questions, showing understanding of the text when referring to explicit details and examples in the text and when explaining inferences drawn from the text.

Writing

Written Expression: students typically address the prompts and provide basic development of ideas, including when drawing evidence from multiple sources, while in the majority of instances demonstrating organization that sometimes is controlled. Students typically:

- Develop topic and/or narrative elements in a manner that is general in its appropriateness to the task and purpose.
- Demonstrate some organization.
- Include some linking words and phrases, descriptive words, and/or temporal words, limiting the clarity with which ideas are expressed.

Knowledge and use of Language Conventions: students typically demonstrate basic command of the conventions of Standard English consistent with edited writing. There are few patterns of errors in grammar and usage that impede understanding, demonstrating partial control over language.

4

To view the full version of the performance level descriptors, visit https://coassessments.com/parentsandguardians.

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

Page 1 of 2

02022022-Z9999999-5555-4444 - 0000000

FIRSTNAME008 (ROBERTA) LASTNAME008

English Language Arts/Literacy

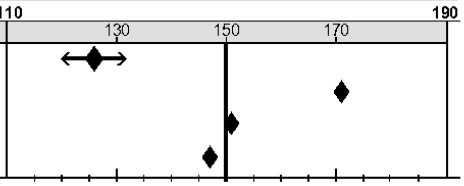
Confidential

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 ares of reading and writing.
District Averages are provided for comparison.
State Averages are provided for comparison.
The dark vertical line indicates the average percent of points earned by students who just crossed into the Met Expectations performance level on the overall English Language Arts/Literacy test.

Reading - Refer to page 1 for participation rates.

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



Reading Scale Score	
Student	126
School	171
District	151
State	147

	Points Possible	Percent of Points Earned* 0% 25% 50% 75% 100%
Literary Text Students read and analyze fiction, drama, and poetry.	18	67%
Informational Text Students read and analyze nonfiction, history, science, and the arts.	18	0%
Vocabulary Students use context to determine what words and phrases mean.	10	20%

	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.	27	100%
Written Expression Students compose well-developed writing using details from what they have read.	7	100%
Language and Conventions Students demonstrate knowledge of conventions and other important elements of language.	6	100%

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

For more information about the standards included in this assessment, please visit the Colorado Department of Education's website at http://www.cde.state.co.us/coreadingwriting/statestandards.

Page 2 of 2

Confidential Student Performance Report

Colorado Measures of Academic Success

Student: **FIRSTNAME M. (SUZIEQ)**
LASTNAMEWWWWWWWWWWWWWWWW

SASID: **9999999999** Birthdate: **MM/DD/CCYY**
School: **SCHOOL NAME (9999)**
District: **DISTRICT NAME (9999)**

Spring 2022

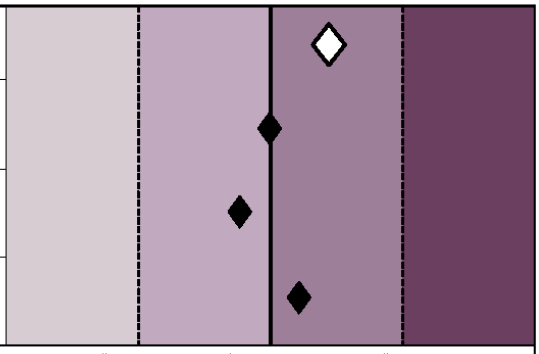
Science — Grade 5

The 2020 Colorado Academic Standards in science were tested for the first time through the spring 2022 Colorado Measures of Academic Success (CMAS). This report provides information about how your student performed on that test in comparison to other Colorado students who took the test.

- Your student's percentile rank is represented by the white diamond on the graph below and indicates the percentage of students whom your student performed higher than in the state. For example, a student in the 37th percentile performed higher than 37% of the students who tested in the state.
- You may compare your student's performance to the performance of other groups of students. State, district and school average percentile ranks are represented by black diamonds on the graph. To protect student privacy, scores are not displayed for districts and schools that do not meet the minimum number of students for reporting. The solid line on the graph indicates the 50th percentile and represents the state average.

Percentile ranks can change based on the group of students tested. Higher participation rates (the percentage of students who took the test) allow for stronger comparisons to all students in the state, district and school. Participation rates are included in the first column in the graph below.

Your student's percentile rank:	**XXth**
Average state percentile rank: (State participation rate: xx%)	50th
Average district percentile rank: (District participation rate: xx%)	XXth
Average school percentile rank: (School participation rate: xx%)	XXth



This report provides information on your student's performance in relation to other students' performance. Unlike other CMAS reports, it does not provide information on your student's level of mastery of the concepts and skills in the Colorado Academic Standards. Performance levels will be determined in the future, in part based on student performance on this 2022 test.

5

For more information about the standards included in this assessment, please visit the Colorado Department of Education's website at http://www.cde.state.co.us/coscience/statestandards.

mmddyyy-Batch-9999-9999-1234567

Appendix E: Student Participation by Demographic Group

Table E.1. Student Participation N-Count Demographic Distribution—Mathematics

Subgroup	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
No IEP	49,833	50,155	50,481	49,513	49,285	47,307
IEP	6,649	6,731	6,942	6,423	5,998	5,497
No Accommodation	43,595	44,581	45,866	48,943	48,669	46,893
Accommodation	12,887	12,305	11,557	6,993	6,614	5,911
Am. Indian/Alaska Native	359	364	347	359	366	312
Asian	1,973	1,898	1,957	1,740	1,766	1,705
Black	2,572	2,507	2,509	2,489	2,419	2,292
Hispanic	18,976	19,428	20,045	19,998	20,283	19,685
White	29,415	29,601	29,446	28,466	27,744	26,262
Hawaiian/Pacific Islander	185	164	181	142	159	155
Two or More Races	2,995	2,919	2,934	2,736	2,539	2,386
Missing	*	*	*	*	*	*
No Economic Disadvantage	19,363	19,610	19,557	19,242	18,692	17,841
Economic Disadvantage	37,119	37,276	37,866	36,694	36,591	34,963
Female	27,925	27,698	28,087	27,334	26,597	25,123
Male	28,557	29,188	29,336	28,602	28,686	27,681
Language Proficiency NA	46,647	46,588	46,647	45,325	44,726	43,528
Language Proficiency NEP	2,619	2,005	1,325	1,111	1,219	1,169
Language Proficiency LEP	6,181	6,245	5,756	4,366	4,020	3,940
Language Proficiency FEP	1,035	2,048	3,695	5,134	5,318	4,167
Not Migrant	56,310	56,733	57,252	55,775	55,120	52,646
Migrant	172	153	171	161	163	158

*n-count less than 16

Table E.2. Student Participation N-Count Demographic Distribution—ELA

Subgroup	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
No IEP	48,587	49,108	50,416	49,488	49,224	47,217
IEP	6,494	6,638	6,946	6,472	6,041	5,510
No Accommodation	51,253	51,420	52,562	51,549	50,867	48,507
Accommodation	3,828	4,326	4,800	4,411	4,398	4,220
Am. Indian/Alaska Native	364	364	352	363	365	316
Asian	1,955	1,887	1,933	1,730	1,751	1,708
Black	2,581	2,513	2,516	2,508	2,437	2,291
Hispanic	17,626	18,286	19,958	19,956	20,235	19,603
White	29,363	29,602	29,480	28,507	27,769	26,270
Hawaiian/Pacific Islander	186	166	183	144	163	158
Two or More Races	2,999	2,922	2,937	2,746	2,538	2,375
Missing	*	*	*	*	*	*
No Economic Disadvantage	36,667	36,866	37,825	36,662	36,531	34,904
Economic Disadvantage	18,414	18,880	19,537	19,298	18,734	17,823
Female	27,216	27,174	28,094	27,353	26,616	25,104
Male	27,865	28,572	29,268	28,607	28,649	27,623
Language Proficiency NA	46,663	46,657	46,746	45,488	44,856	43,574
Language Proficiency NEP	1,956	1,466	1,130	928	1,052	1,032
Language Proficiency LEP	5,426	5,570	5,783	4,377	4,024	3,950
Language Proficiency FEP	1,036	2,053	3,703	5,167	5,333	4,171
Not Migrant	54,916	55,592	57,192	55,799	55,104	52,569
Migrant	165	154	170	161	161	158

*n-count less than 16

Table E.3. Student Participation N-Count Demographic Distribution—CSLA

Subgroup	Grade 3	Grade 4
No IEP	1,185	1,059
IEP	116	92
No Accommodation	1,151	986
Accommodation	150	165
Am. Indian/Alaska Native	*	*
Asian	*	*
Black	*	*
Hispanic	1,291	1,140
White	*	*
Hawaiian/Pacific Islander	*	*
Two or More Races	*	*
Missing	*	*
No Economic Disadvantage	394	404
Economic Disadvantage	907	747
Female	685	544
Male	616	607
Language Proficiency NA	*	*
Language Proficiency NEP	558	466
Language Proficiency LEP	743	685
Language Proficiency FEP	*	*
Not Migrant	1,295	1,150
Migrant	*	*

*n-count less than 16

Table E.4. Student Participation N-Count Demographic Distribution—Science

Subgroup	Grade 5	Grade 8	Grade 11
No IEP	49,984	46,187	27,992
IEP	6,836	5,340	2,632
No Accommodation	53,900	48,974	29,682
Accommodation	2,920	2,553	942
Am. Indian/Alaska Native	338	306	210
Asian	1,945	1,679	931
Black	2,494	2,209	1,374
Hispanic	19,822	19,209	12,585
White	29,136	25,668	14,207
Hawaiian/Pacific Islander	176	145	89
Two or More Races	2,905	2,305	1,227
Missing	*	*	*
No Economic Disadvantage	37,487	34,124	20,561
Economic Disadvantage	19,333	17,403	10,063
Female	27,818	24,468	14,418
Male	29,002	27,059	16,206
Language Proficiency NA	46,133	42,480	26,370
Language Proficiency NEP	1,296	1,132	592
Language Proficiency LEP	5,713	3,838	1,905
Language Proficiency FEP	3,678	4,077	1,757
Not Migrant	56,654	51,371	30,536
Migrant	166	156	88

*n-count less than 16

Appendix F: Scale Score Distributions

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	857	1.52	857	1.52
651	46	0.08	903	1.60
652	72	0.13	975	1.73
653	62	0.11	1037	1.84
654	49	0.09	1086	1.92
655	68	0.12	1154	2.04
656	84	0.15	1238	2.19
657	77	0.14	1315	2.33
658	68	0.12	1383	2.45
659	110	0.19	1493	2.64
660	95	0.17	1588	2.81
661	95	0.17	1683	2.98
662	115	0.20	1798	3.18
663	104	0.18	1902	3.37
664	117	0.21	2019	3.57
665	137	0.24	2156	3.82
666	126	0.22	2282	4.04
667	137	0.24	2419	4.28
668	139	0.25	2558	4.53
669	144	0.25	2702	4.78
670	153	0.27	2855	5.05
671	150	0.27	3005	5.32
672	157	0.28	3162	5.60
673	161	0.29	3323	5.88
674	190	0.34	3513	6.22
675	186	0.33	3699	6.55
676	198	0.35	3897	6.90
677	192	0.34	4089	7.24
678	202	0.36	4291	7.60
679	191	0.34	4482	7.94
680	233	0.41	4715	8.35
681	266	0.47	4981	8.82
682	249	0.44	5230	9.26
683	281	0.50	5511	9.76
684	255	0.45	5766	10.21
685	258	0.46	6024	10.67
686	273	0.48	6297	11.15
687	328	0.58	6625	11.73
688	273	0.48	6898	12.21
689	318	0.56	7216	12.78
690	285	0.50	7501	13.28
691	317	0.56	7818	13.84
692	335	0.59	8153	14.43
693	328	0.58	8481	15.02
694	312	0.55	8793	15.57
695	356	0.63	9149	16.20
696	341	0.60	9490	16.80
697	350	0.62	9840	17.42
698	363	0.64	10203	18.06
699	346	0.61	10549	18.68
700	355	0.63	10904	19.31
701	348	0.62	11252	19.92
702	374	0.66	11626	20.58
703	360	0.64	11986	21.22
704	374	0.66	12360	21.88
705	372	0.66	12732	22.54
706	411	0.73	13143	23.27
707	429	0.76	13572	24.03
708	386	0.68	13958	24.71
709	425	0.75	14383	25.46
710	417	0.74	14800	26.20
711	431	0.76	15231	26.97
712	450	0.80	15681	27.76
713	400	0.71	16081	28.47
714	435	0.77	16516	29.24
715	478	0.85	16994	30.09
716	435	0.77	17429	30.86
717	442	0.78	17871	31.64
718	504	0.89	18375	32.53
719	467	0.83	18842	33.36
720	474	0.84	19316	34.20
721	475	0.84	19791	35.04
722	489	0.87	20280	35.91
723	491	0.87	20771	36.77
724	484	0.86	21255	37.63
725	498	0.88	21753	38.51
726	490	0.87	22243	39.38
727	478	0.85	22721	40.23
728	520	0.92	23241	41.15
729	480	0.85	23721	42.00
730	505	0.89	24226	42.89
731	544	0.96	24770	43.85
732	546	0.97	25316	44.82
733	524	0.93	25840	45.75
734	492	0.87	26332	46.62
735	518	0.92	26850	47.54
736	523	0.93	27373	48.46
737	577	1.02	27950	49.48
738	484	0.86	28434	50.34
739	525	0.93	28959	51.27
740	512	0.91	29471	52.18
741	538	0.95	30009	53.13
742	513	0.91	30522	54.04
743	550	0.97	31072	55.01
744	535	0.95	31607	55.96
745	536	0.95	32143	56.91

SS	Freq.	%	Cum. Freq.	Cum. %
746	512	0.91	32655	57.81
747	553	0.98	33208	58.79
748	522	0.92	33730	59.72
749	515	0.91	34245	60.63
750	542	0.96	34787	61.59
751	525	0.93	35312	62.52
752	514	0.91	35826	63.43
753	548	0.97	36374	64.40
754	536	0.95	36910	65.35
755	496	0.88	37406	66.23
756	521	0.92	37927	67.15
757	529	0.94	38456	68.09
758	555	0.98	39011	69.07
759	531	0.94	39542	70.01
760	496	0.88	40038	70.89
761	492	0.87	40530	71.76
762	520	0.92	41050	72.68
763	492	0.87	41542	73.55
764	501	0.89	42043	74.44
765	505	0.89	42548	75.33
766	463	0.82	43011	76.15
767	473	0.84	43484	76.99
768	453	0.80	43937	77.79
769	449	0.79	44386	78.58
770	458	0.81	44844	79.40
771	451	0.80	45295	80.19
772	424	0.75	45719	80.94
773	448	0.79	46167	81.74
774	433	0.77	46600	82.50
775	381	0.67	46981	83.18
776	400	0.71	47381	83.89
777	363	0.64	47744	84.53
778	342	0.61	48086	85.14
779	372	0.66	48458	85.79
780	350	0.62	48808	86.41
781	322	0.57	49130	86.98
782	347	0.61	49477	87.60
783	350	0.62	49827	88.22
784	323	0.57	50150	88.79
785	303	0.54	50453	89.33
786	298	0.53	50751	89.85
787	288	0.51	51039	90.36
788	282	0.50	51321	90.86
789	279	0.49	51600	91.36
790	277	0.49	51877	91.85
791	254	0.45	52131	92.30
792	224	0.40	52355	92.69
793	232	0.41	52587	93.10
794	224	0.40	52811	93.50
795	169	0.30	52980	93.80
796	190	0.34	53170	94.14
797	188	0.33	53358	94.47
798	151	0.27	53509	94.74
799	149	0.26	53658	95.00
800	154	0.27	53812	95.27
801	163	0.29	53975	95.56
802	134	0.24	54109	95.80
803	130	0.23	54239	96.03
804	140	0.25	54379	96.28
805	107	0.19	54486	96.47
806	114	0.20	54600	96.67
807	91	0.16	54691	96.83
808	108	0.19	54799	97.02
809	96	0.17	54895	97.19
810	116	0.21	55011	97.40
811	87	0.15	55098	97.55
812	109	0.19	55207	97.74
813	62	0.11	55269	97.85
814	84	0.15	55353	98.00
815	55	0.10	55408	98.10
816	82	0.15	55490	98.24
817	64	0.11	55554	98.36
818	55	0.10	55609	98.45
819	72	0.13	55681	98.58
820	32	0.06	55713	98.64
821	56	0.10	55769	98.74
822	32	0.06	55801	98.79
823	29	0.05	55830	98.85
824	48	0.08	55878	98.93
825	50	0.09	55928	99.02
826	30	0.05	55958	99.07
827	41	0.07	55999	99.14
828	36	0.06	56035	99.21
829	19	0.03	56054	99.24
830	32	0.06	56086	99.30
831	31	0.05	56117	99.35
832	13	0.02	56130	99.38
833	13	0.02	56143	99.40
834	22	0.04	56165	99.44
835	40	0.07	56205	99.51
836	18	0.03	56223	99.54
837	8	0.01	56231	99.56
838	22	0.04	56253	99.59
839	13	0.02	56266	99.62
840	9	0.02	56275	99.63
841	5	0.01	56280	99.64
842	10	0.02	56290	99.66
843	21	0.04	56311	99.70
844	20	0.04	56331	99.73
845	4	0.01	56335	99.74
846	3	0.01	56338	99.75
847	11	0.02	56349	99.76
848	20	0.04	56369	99.80
849	8	0.01	56377	99.81
850	105	0.19	56482	100.00

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	304	0.53	304	0.53
651	38	0.07	342	0.60
652	14	0.02	356	0.63
653	6	0.01	362	0.64
654	38	0.07	400	0.70
655	48	0.08	448	0.79
656	22	0.04	470	0.83
657	25	0.04	495	0.87
658	21	0.04	516	0.91
659	25	0.04	541	0.95
660	77	0.14	618	1.09
661	34	0.06	652	1.15
662	35	0.06	687	1.21
663	107	0.19	794	1.40
664	64	0.11	858	1.51
665	57	0.10	915	1.61
666	80	0.14	995	1.75
667	78	0.14	1073	1.89
668	58	0.10	1131	1.99
669	109	0.19	1240	2.18
670	58	0.10	1298	2.28
671	131	0.23	1429	2.51
672	142	0.25	1571	2.76
673	111	0.20	1682	2.96
674	180	0.32	1862	3.27
675	126	0.22	1988	3.49
676	194	0.34	2182	3.84
677	191	0.34	2373	4.17
678	166	0.29	2539	4.46
679	176	0.31	2715	4.77
680	233	0.41	2948	5.18
681	199	0.35	3147	5.53
682	241	0.42	3388	5.96
683	263	0.46	3651	6.42
684	291	0.51	3942	6.93
685	309	0.54	4251	7.47
686	322	0.57	4573	8.04
687	324	0.57	4897	8.61
688	375	0.66	5272	9.27
689	391	0.69	5663	9.95
690	388	0.68	6051	10.64
691	388	0.68	6439	11.32
692	412	0.72	6851	12.04
693	455	0.80	7306	12.84
694	444	0.78	7750	13.62
695	481	0.85	8231	14.47
696	521	0.92	8752	15.39
697	509	0.89	9261	16.28
698	468	0.82	9729	17.10
699	537	0.94	10266	18.05
700	526	0.92	10792	18.97
701	504	0.89	11296	19.86
702	529	0.93	11825	20.79
703	558	0.98	12383	21.77
704	551	0.97	12934	22.74
705	575	1.01	13509	23.75
706	551	0.97	14060	24.72
707	511	0.90	14571	25.61
708	556	0.98	15127	26.59
709	537	0.94	15664	27.54
710	540	0.95	16204	28.49
711	558	0.98	16762	29.47
712	565	0.99	17327	30.46
713	517	0.91	17844	31.37
714	576	1.01	18420	32.38
715	566	0.99	18986	33.38
716	565	0.99	19551	34.37
717	520	0.91	20071	35.28
718	549	0.97	20620	36.25
719	540	0.95	21160	37.20
720	543	0.95	21703	38.15
721	539	0.95	22242	39.10
722	560	0.98	22802	40.08
723	583	1.02	23385	41.11
724	562	0.99	23947	42.10
725	572	1.01	24519	43.10
726	587	1.03	25106	44.13
727	592	1.04	25698	45.17
728	613	1.08	26311	46.25
729	628	1.10	26939	47.36
730	602	1.06	27541	48.41
731	620	1.09	28161	49.50
732	609	1.07	28770	50.57
733	617	1.08	29387	51.66
734	635	1.12	30022	52.78
735	640	1.13	30662	53.90
736	636	1.12	31298	55.02
737	647	1.14	31945	56.16
738	635	1.12	32580	57.27
739	620	1.09	33200	58.36
740	638	1.12	33838	59.48
741	626	1.10	34464	60.58
742	667	1.17	35131	61.76
743	619	1.09	35750	62.84
744	637	1.12	36387	63.96
745	632	1.11	37019	65.08
746	584	1.03	37603	66.10
747	602	1.06	38205	67.16
748	618	1.09	38823	68.25
749	596	1.05	39419	69.29

SS	Freq.	%	Cum. Freq.	Cum. %
750	585	1.03	40004	70.32
751	585	1.03	40589	71.35
752	600	1.05	41189	72.41
753	610	1.07	41799	73.48
754	582	1.02	42381	74.50
755	595	1.05	42976	75.55
756	534	0.94	43510	76.49
757	526	0.92	44036	77.41
758	497	0.87	44533	78.28
759	508	0.89	45041	79.18
760	471	0.83	45512	80.01
761	511	0.90	46023	80.90
762	499	0.88	46522	81.78
763	459	0.81	46981	82.59
764	432	0.76	47413	83.35
765	451	0.79	47864	84.14
766	375	0.66	48239	84.80
767	432	0.76	48671	85.56
768	423	0.74	49094	86.30
769	399	0.70	49493	87.00
770	392	0.69	49885	87.69
771	353	0.62	50238	88.31
772	354	0.62	50592	88.94
773	343	0.60	50935	89.54
774	329	0.58	51264	90.12
775	327	0.57	51591	90.69
776	300	0.53	51891	91.22
777	321	0.56	52212	91.78
778	285	0.50	52497	92.28
779	277	0.49	52774	92.77
780	248	0.44	53022	93.21
781	219	0.38	53241	93.59
782	218	0.38	53459	93.98
783	203	0.36	53662	94.33
784	230	0.40	53892	94.74
785	172	0.30	54064	95.04
786	178	0.31	54242	95.35
787	182	0.32	54424	95.67
788	189	0.33	54613	96.00
789	161	0.28	54774	96.29
790	133	0.23	54907	96.52
791	137	0.24	55044	96.76
792	141	0.25	55185	97.01
793	122	0.21	55307	97.22
794	100	0.18	55407	97.40
795	99	0.17	55506	97.57
796	101	0.18	55607	97.75
797	83	0.15	55690	97.90
798	94	0.17	55784	98.06
799	81	0.14	55865	98.21
800	78	0.14	55943	98.34
801	72	0.13	56015	98.47
802	59	0.10	56074	98.57
803	49	0.09	56123	98.66
804	53	0.09	56176	98.75
805	50	0.09	56226	98.84
806	52	0.09	56278	98.93
807	44	0.08	56322	99.01
808	40	0.07	56362	99.08
809	22	0.04	56384	99.12
810	29	0.05	56413	99.17
811	32	0.06	56445	99.22
812	32	0.06	56477	99.28
813	26	0.05	56503	99.33
814	28	0.05	56531	99.38
815	18	0.03	56549	99.41
816	21	0.04	56570	99.44
817	27	0.05	56597	99.49
818	21	0.04	56618	99.53
819	12	0.02	56630	99.55
820	11	0.02	56641	99.57
821	13	0.02	56654	99.59
822	9	0.02	56663	99.61
823	11	0.02	56674	99.63
824	14	0.02	56688	99.65
825	15	0.03	56703	99.68
826	5	0.01	56708	99.69
827	19	0.03	56727	99.72
828	7	0.01	56734	99.73
829	11	0.02	56745	99.75
830	5	0.01	56750	99.76
831	6	0.01	56756	99.77
832	14	0.02	56770	99.80
833	12	0.02	56782	99.82
834	11	0.02	56793	99.84
835	1	0.00	56794	99.84
836	4	0.01	56798	99.85
837	6	0.01	56804	99.86
839	9	0.02	56813	99.87
842	3	0.01	56816	99.88
844	2	0.00	56818	99.88
846	2	0.00	56820	99.88
847	2	0.00	56822	99.89
849	5	0.01	56827	99.90
850	59	0.10	56886	100.00

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	236	0.41	236	0.41
651	2	0.00	238	0.41
652	52	0.09	290	0.51
653	22	0.04	312	0.54
654	24	0.04	336	0.59
655	66	0.11	402	0.70
656	72	0.13	474	0.83
657	36	0.06	510	0.89
658	24	0.04	534	0.93
659	50	0.09	584	1.02
660	24	0.04	608	1.06
661	32	0.06	640	1.11
662	42	0.07	682	1.19
663	11	0.02	693	1.21
664	38	0.07	731	1.27
665	31	0.05	762	1.33
666	74	0.13	836	1.46
667	97	0.17	933	1.62
668	84	0.15	1017	1.77
669	128	0.22	1145	1.99
670	90	0.16	1235	2.15
671	89	0.15	1324	2.31
672	91	0.16	1415	2.46
673	84	0.15	1499	2.61
674	109	0.19	1608	2.80
675	107	0.19	1715	2.99
676	146	0.25	1861	3.24
677	179	0.31	2040	3.55
678	206	0.36	2246	3.91
679	156	0.27	2402	4.18
680	181	0.32	2583	4.50
681	171	0.30	2754	4.80
682	223	0.39	2977	5.18
683	247	0.43	3224	5.61
684	276	0.48	3500	6.10
685	261	0.45	3761	6.55
686	262	0.46	4023	7.01
687	285	0.50	4308	7.50
688	320	0.56	4628	8.06
689	372	0.65	5000	8.71
690	380	0.66	5380	9.37
691	382	0.67	5762	10.03
692	370	0.64	6132	10.68
693	407	0.71	6539	11.39
694	405	0.71	6944	12.09
695	441	0.77	7385	12.86
696	432	0.75	7817	13.61
697	496	0.86	8313	14.48
698	473	0.82	8786	15.30
699	502	0.87	9288	16.17
700	499	0.87	9787	17.04
701	506	0.88	10293	17.92
702	497	0.87	10790	18.79
703	520	0.91	11310	19.70
704	541	0.94	11851	20.64
705	565	0.98	12416	21.62
706	500	0.87	12916	22.49
707	561	0.98	13477	23.47
708	535	0.93	14012	24.40
709	548	0.95	14560	25.36
710	540	0.94	15100	26.30
711	558	0.97	15658	27.27
712	603	1.05	16261	28.32
713	607	1.06	16868	29.37
714	586	1.02	17454	30.40
715	571	0.99	18025	31.39
716	557	0.97	18582	32.36
717	540	0.94	19122	33.30
718	645	1.12	19767	34.42
719	596	1.04	20363	35.46
720	555	0.97	20918	36.43
721	605	1.05	21523	37.48
722	594	1.03	22117	38.52
723	565	0.98	22682	39.50
724	565	0.98	23247	40.48
725	564	0.98	23811	41.47
726	591	1.03	24402	42.50
727	561	0.98	24963	43.47
728	588	1.02	25551	44.50
729	574	1.00	26125	45.50
730	554	0.96	26679	46.46
731	548	0.95	27227	47.41
732	587	1.02	27814	48.44
733	583	1.02	28397	49.45
734	575	1.00	28972	50.45
735	584	1.02	29556	51.47
736	598	1.04	30154	52.51
737	548	0.95	30702	53.47
738	605	1.05	31307	54.52
739	564	0.98	31871	55.50
740	593	1.03	32464	56.53
741	612	1.07	33076	57.60
742	551	0.96	33627	58.56
743	515	0.90	34142	59.46
744	545	0.95	34687	60.41
745	557	0.97	35244	61.38
746	534	0.93	35778	62.31
747	528	0.92	36306	63.23
748	517	0.90	36823	64.13
749	544	0.95	37367	65.07

SS	Freq.	%	Cum. Freq.	Cum. %
750	515	0.90	37882	65.97
751	519	0.90	38401	66.87
752	544	0.95	38945	67.82
753	520	0.91	39465	68.73
754	476	0.83	39941	69.56
755	521	0.91	40462	70.46
756	478	0.83	40940	71.30
757	478	0.83	41418	72.13
758	538	0.94	41956	73.06
759	489	0.85	42445	73.92
760	494	0.86	42939	74.78
761	453	0.79	43392	75.57
762	498	0.87	43890	76.43
763	453	0.79	44343	77.22
764	453	0.79	44796	78.01
765	424	0.74	45220	78.75
766	458	0.80	45678	79.55
767	405	0.71	46083	80.25
768	437	0.76	46520	81.01
769	395	0.69	46915	81.70
770	400	0.70	47315	82.40
771	402	0.70	47717	83.10
772	407	0.71	48124	83.81
773	336	0.59	48460	84.39
774	370	0.64	48830	85.04
775	350	0.61	49180	85.65
776	366	0.64	49546	86.28
777	339	0.59	49885	86.87
778	343	0.60	50228	87.47
779	350	0.61	50578	88.08
780	333	0.58	50911	88.66
781	317	0.55	51228	89.21
782	313	0.55	51541	89.76
783	309	0.54	51850	90.29
784	287	0.50	52137	90.79
785	301	0.52	52438	91.32
786	293	0.51	52731	91.83
787	281	0.49	53012	92.32
788	241	0.42	53253	92.74
789	229	0.40	53482	93.14
790	245	0.43	53727	93.56
791	245	0.43	53972	93.99
792	206	0.36	54178	94.35
793	225	0.39	54403	94.74
794	192	0.33	54595	95.08
795	170	0.30	54765	95.37
796	191	0.33	54956	95.70
797	144	0.25	55100	95.95
798	155	0.27	55255	96.22
799	144	0.25	55399	96.48
800	149	0.26	55548	96.73
801	131	0.23	55679	96.96
802	120	0.21	55799	97.17
803	129	0.22	55928	97.40
804	102	0.18	56030	97.57
805	93	0.16	56123	97.74
806	96	0.17	56219	97.90
807	95	0.17	56314	98.07
808	77	0.13	56391	98.20
809	86	0.15	56477	98.35
810	69	0.12	56546	98.47
811	71	0.12	56617	98.60
812	67	0.12	56684	98.71
813	51	0.09	56735	98.80
814	68	0.12	56803	98.92
815	41	0.07	56844	98.99
816	41	0.07	56885	99.06
817	43	0.07	56928	99.14
818	36	0.06	56964	99.20
819	42	0.07	57006	99.27
820	22	0.04	57028	99.31
821	36	0.06	57064	99.37
822	20	0.03	57084	99.41
823	34	0.06	57118	99.47
824	32	0.06	57150	99.52
825	24	0.04	57174	99.57
826	18	0.03	57192	99.60
827	12	0.02	57204	99.62
828	18	0.03	57222	99.65
829	18	0.03	57240	99.68
830	5	0.01	57245	99.69
831	11	0.02	57256	99.71
832	18	0.03	57274	99.74
833	5	0.01	57279	99.75
834	18	0.03	57297	99.78
835	9	0.02	57306	99.80
836	13	0.02	57319	99.82
837	4	0.01	57323	99.83
838	15	0.03	57338	99.85
839	9	0.02	57347	99.87
840	1	0.00	57348	99.87
841	3	0.01	57351	99.87
842	5	0.01	57356	99.88
843	2	0.00	57358	99.89
844	4	0.01	57362	99.89
850	61	0.11	57423	100.00

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	483	0.86	483	0.86
651	20	0.04	503	0.90
652	8	0.01	511	0.91
653	27	0.05	538	0.96
654	5	0.01	543	0.97
655	123	0.22	666	1.19
656	34	0.06	700	1.25
657	23	0.04	723	1.29
658	45	0.08	768	1.37
659	26	0.05	794	1.42
660	24	0.04	818	1.46
661	52	0.09	870	1.56
662	145	0.26	1015	1.81
663	46	0.08	1061	1.90
664	88	0.16	1149	2.05
665	63	0.11	1212	2.17
666	60	0.11	1272	2.27
667	161	0.29	1433	2.56
668	74	0.13	1507	2.69
669	133	0.24	1640	2.93
670	106	0.19	1746	3.12
671	111	0.20	1857	3.32
672	177	0.32	2034	3.64
673	115	0.21	2149	3.84
674	239	0.43	2388	4.27
675	188	0.34	2576	4.61
676	195	0.35	2771	4.95
677	146	0.26	2917	5.21
678	296	0.53	3213	5.74
679	220	0.39	3433	6.14
680	189	0.34	3622	6.48
681	299	0.53	3921	7.01
682	255	0.46	4176	7.47
683	306	0.55	4482	8.01
684	336	0.60	4818	8.61
685	308	0.55	5126	9.16
686	345	0.62	5471	9.78
687	356	0.64	5827	10.42
688	429	0.77	6256	11.18
689	406	0.73	6662	11.91
690	433	0.77	7095	12.68
691	407	0.73	7502	13.41
692	428	0.77	7930	14.18
693	454	0.81	8384	14.99
694	494	0.88	8878	15.87
695	467	0.83	9345	16.71
696	507	0.91	9852	17.61
697	478	0.85	10330	18.47
698	510	0.91	10840	19.38
699	530	0.95	11370	20.33
700	565	1.01	11935	21.34
701	539	0.96	12474	22.30
702	567	1.01	13041	23.31
703	539	0.96	13580	24.28
704	560	1.00	14140	25.28
705	580	1.04	14720	26.32
706	557	1.00	15277	27.31
707	565	1.01	15842	28.32
708	573	1.02	16415	29.35
709	552	0.99	16967	30.33
710	544	0.97	17511	31.31
711	549	0.98	18060	32.29
712	581	1.04	18641	33.33
713	571	1.02	19212	34.35
714	578	1.03	19790	35.38
715	627	1.12	20417	36.50
716	615	1.10	21032	37.60
717	626	1.12	21658	38.72
718	618	1.10	22276	39.82
719	649	1.16	22925	40.98
720	595	1.06	23520	42.05
721	593	1.06	24113	43.11
722	594	1.06	24707	44.17
723	587	1.05	25294	45.22
724	632	1.13	25926	46.35
725	623	1.11	26549	47.46
726	641	1.15	27190	48.61
727	600	1.07	27790	49.68
728	629	1.12	28419	50.81
729	613	1.10	29032	51.90
730	629	1.12	29661	53.03
731	616	1.10	30277	54.13
732	638	1.14	30915	55.27
733	636	1.14	31551	56.41
734	611	1.09	32162	57.50
735	605	1.08	32767	58.58
736	644	1.15	33411	59.73
737	628	1.12	34039	60.85
738	592	1.06	34631	61.91
739	637	1.14	35268	63.05
740	602	1.08	35870	64.13
741	611	1.09	36481	65.22
742	604	1.08	37085	66.30
743	623	1.11	37708	67.41
744	590	1.05	38298	68.47
745	603	1.08	38901	69.55
746	554	0.99	39455	70.54
747	579	1.04	40034	71.57
748	598	1.07	40632	72.64
749	604	1.08	41236	73.72

SS	Freq.	%	Cum. Freq.	Cum. %
750	604	1.08	41840	74.80
751	558	1.00	42398	75.80
752	527	0.94	42925	76.74
753	560	1.00	43485	77.74
754	545	0.97	44030	78.71
755	539	0.96	44569	79.68
756	524	0.94	45093	80.62
757	495	0.88	45588	81.50
758	484	0.87	46072	82.37
759	430	0.77	46502	83.13
760	411	0.73	46913	83.87
761	434	0.78	47347	84.64
762	407	0.73	47754	85.37
763	421	0.75	48175	86.13
764	384	0.69	48559	86.81
765	345	0.62	48904	87.43
766	361	0.65	49265	88.07
767	323	0.58	49588	88.65
768	308	0.55	49896	89.20
769	312	0.56	50208	89.76
770	312	0.56	50520	90.32
771	294	0.53	50814	90.84
772	276	0.49	51090	91.34
773	257	0.46	51347	91.80
774	266	0.48	51613	92.27
775	238	0.43	51851	92.70
776	239	0.43	52090	93.12
777	214	0.38	52304	93.51
778	221	0.40	52525	93.90
779	192	0.34	52717	94.25
780	214	0.38	52931	94.63
781	190	0.34	53121	94.97
782	176	0.31	53297	95.28
783	158	0.28	53455	95.56
784	166	0.30	53621	95.86
785	125	0.22	53746	96.08
786	148	0.26	53894	96.35
787	118	0.21	54012	96.56
788	125	0.22	54137	96.78
789	110	0.20	54247	96.98
790	107	0.19	54354	97.17
791	94	0.17	54448	97.34
792	99	0.18	54547	97.52
793	102	0.18	54649	97.70
794	88	0.16	54737	97.86
795	82	0.15	54819	98.00
796	71	0.13	54890	98.13
797	73	0.13	54963	98.26
798	64	0.11	55027	98.37
799	63	0.11	55090	98.49
800	53	0.09	55143	98.58
801	48	0.09	55191	98.67
802	63	0.11	55254	98.78
803	40	0.07	55294	98.85
804	55	0.10	55349	98.95
805	42	0.08	55391	99.03
806	32	0.06	55423	99.08
807	36	0.06	55459	99.15
808	27	0.05	55486	99.20
809	35	0.06	55521	99.26
810	32	0.06	55553	99.32
811	24	0.04	55577	99.36
812	23	0.04	55600	99.40
813	20	0.04	55620	99.44
814	28	0.05	55648	99.49
815	13	0.02	55661	99.51
816	26	0.05	55687	99.55
817	26	0.05	55713	99.60
818	14	0.03	55727	99.63
819	19	0.03	55746	99.66
820	10	0.02	55756	99.68
821	15	0.03	55771	99.71
822	11	0.02	55782	99.72
823	3	0.01	55785	99.73
824	9	0.02	55794	99.75
825	5	0.01	55799	99.76
826	9	0.02	55808	99.77
827	8	0.01	55816	99.79
828	11	0.02	55827	99.81
829	3	0.01	55830	99.81
830	12	0.02	55842	99.83
831	1	0.00	55843	99.83
832	7	0.01	55850	99.85
833	2	0.00	55852	99.85
834	5	0.01	55857	99.86
835	5	0.01	55862	99.87
836	4	0.01	55866	99.87
837	8	0.01	55874	99.89
838	2	0.00	55876	99.89
839	2	0.00	55878	99.90
840	1	0.00	55879	99.90
841	3	0.01	55882	99.90
842	5	0.01	55887	99.91
843	8	0.01	55895	99.93
845	2	0.00	55897	99.93
846	3	0.01	55900	99.94
847	3	0.01	55903	99.94
849	1	0.00	55904	99.94
850	32	0.06	55936	100.00

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	172	0.31	172	0.31
651	21	0.04	193	0.35
654	27	0.05	220	0.40
655	25	0.05	245	0.44
656	6	0.01	251	0.45
657	24	0.04	275	0.50
658	1	0.00	276	0.50
659	33	0.06	309	0.56
660	37	0.07	346	0.63
661	5	0.01	351	0.63
662	38	0.07	389	0.70
663	6	0.01	395	0.71
664	15	0.03	410	0.74
665	24	0.04	434	0.79
666	28	0.05	462	0.84
667	34	0.06	496	0.90
668	22	0.04	518	0.94
669	40	0.07	558	1.01
670	35	0.06	593	1.07
671	50	0.09	643	1.16
672	73	0.13	716	1.30
673	92	0.17	808	1.46
674	45	0.08	853	1.54
675	71	0.13	924	1.67
676	84	0.15	1008	1.82
677	67	0.12	1075	1.94
678	78	0.14	1153	2.09
679	114	0.21	1267	2.29
680	101	0.18	1368	2.47
681	106	0.19	1474	2.67
682	117	0.21	1591	2.88
683	143	0.26	1734	3.14
684	138	0.25	1872	3.39
685	167	0.30	2039	3.69
686	194	0.35	2233	4.04
687	245	0.44	2478	4.48
688	229	0.41	2707	4.90
689	224	0.41	2931	5.30
690	280	0.51	3211	5.81
691	275	0.50	3486	6.31
692	310	0.56	3796	6.87
693	306	0.55	4102	7.42
694	331	0.60	4433	8.02
695	422	0.76	4855	8.78
696	392	0.71	5247	9.49
650	172	0.31	172	0.31
651	21	0.04	193	0.35
697	439	0.79	5686	10.29
698	470	0.85	6156	11.14
699	482	0.87	6638	12.01
700	542	0.98	7180	12.99
701	565	1.02	7745	14.01
702	580	1.05	8325	15.06
703	612	1.11	8937	16.17
704	631	1.14	9568	17.31
705	688	1.24	10256	18.55
706	701	1.27	10957	19.82
707	713	1.29	11670	21.11
708	745	1.35	12415	22.46
709	709	1.28	13124	23.74
710	776	1.40	13900	25.14
711	809	1.46	14709	26.61
712	734	1.33	15443	27.93
713	801	1.45	16244	29.38
714	772	1.40	17016	30.78
715	793	1.43	17809	32.21
716	804	1.45	18613	33.67
717	801	1.45	19414	35.12
718	741	1.34	20155	36.46
719	780	1.41	20935	37.87
720	738	1.33	21673	39.20
721	811	1.47	22484	40.67
722	764	1.38	23248	42.05
723	761	1.38	24009	43.43
724	723	1.31	24732	44.74
725	699	1.26	25431	46.00
726	770	1.39	26201	47.39
727	736	1.33	26937	48.73
728	762	1.38	27699	50.10
729	723	1.31	28422	51.41
730	757	1.37	29179	52.78
731	724	1.31	29903	54.09
732	768	1.39	30671	55.48
733	720	1.30	31391	56.78
734	685	1.24	32076	58.02
735	710	1.28	32786	59.31
736	685	1.24	33471	60.54
737	664	1.20	34135	61.75
738	613	1.11	34748	62.85
739	614	1.11	35362	63.97
740	622	1.13	35984	65.09
741	638	1.15	36622	66.24
742	650	1.18	37272	67.42
743	595	1.08	37867	68.50
744	598	1.08	38465	69.58
745	639	1.16	39104	70.73
746	586	1.06	39690	71.79
747	590	1.07	40280	72.86
748	572	1.03	40852	73.90
749	583	1.05	41435	74.95

SS	Freq.	%	Cum. Freq.	Cum. %
750	577	1.04	42012	75.99
751	523	0.95	42535	76.94
752	541	0.98	43076	77.92
753	533	0.96	43609	78.88
754	527	0.95	44136	79.84
755	517	0.94	44653	80.77
756	480	0.87	45133	81.64
757	484	0.88	45617	82.52
758	474	0.86	46091	83.37
759	455	0.82	46546	84.20
760	462	0.84	47008	85.03
761	428	0.77	47436	85.81
762	419	0.76	47855	86.56
763	427	0.77	48282	87.34
764	395	0.71	48677	88.05
765	383	0.69	49060	88.74
766	370	0.67	49430	89.41
767	375	0.68	49805	90.09
768	338	0.61	50143	90.70
769	307	0.56	50450	91.26
770	311	0.56	50761	91.82
771	310	0.56	51071	92.38
772	291	0.53	51362	92.91
773	274	0.50	51636	93.40
774	241	0.44	51877	93.84
775	232	0.42	52109	94.26
776	254	0.46	52363	94.72
777	225	0.41	52588	95.13
778	203	0.37	52791	95.49
779	166	0.30	52957	95.79
780	172	0.31	53129	96.10
781	175	0.32	53304	96.42
782	175	0.32	53479	96.74
783	138	0.25	53617	96.99
784	124	0.22	53741	97.21
785	126	0.23	53867	97.44
786	116	0.21	53983	97.65
787	98	0.18	54081	97.83
788	107	0.19	54188	98.02
789	86	0.16	54274	98.17
790	92	0.17	54366	98.34
791	76	0.14	54442	98.48
792	70	0.13	54512	98.61
793	55	0.10	54567	98.70
794	67	0.12	54634	98.83
795	57	0.10	54691	98.93
796	40	0.07	54731	99.00
797	46	0.08	54777	99.08
798	43	0.08	54820	99.16
799	39	0.07	54859	99.23
800	48	0.09	54907	99.32

SS	Freq.	%	Cum. Freq.	Cum. %
801	40	0.07	54947	99.39
802	27	0.05	54974	99.44
803	37	0.07	55011	99.51
804	24	0.04	55035	99.55
805	17	0.03	55052	99.58
806	12	0.02	55064	99.60
807	18	0.03	55082	99.64
808	16	0.03	55098	99.67
809	17	0.03	55115	99.70
810	11	0.02	55126	99.72
811	19	0.03	55145	99.75
812	18	0.03	55163	99.78
813	11	0.02	55174	99.80
814	8	0.01	55182	99.82
815	6	0.01	55188	99.83
816	7	0.01	55195	99.84
817	6	0.01	55201	99.85
818	10	0.02	55211	99.87
819	4	0.01	55215	99.88
820	2	0.00	55217	99.88
821	4	0.01	55221	99.89
822	6	0.01	55227	99.90
823	12	0.02	55239	99.92
824	5	0.01	55244	99.93
825	1	0.00	55245	99.93
826	4	0.01	55249	99.94
827	3	0.01	55252	99.94
828	1	0.00	55253	99.95
829	2	0.00	55255	99.95
830	3	0.01	55258	99.95
831	1	0.00	55259	99.96
833	4	0.01	55263	99.96
835	1	0.00	55264	99.97
836	1	0.00	55265	99.97
837	1	0.00	55266	99.97
838	1	0.00	55267	99.97
839	2	0.00	55269	99.97
840	3	0.01	55272	99.98
841	1	0.00	55273	99.98
845	4	0.01	55277	99.99
846	1	0.00	55278	99.99
850	5	0.01	55283	100.00

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	653	1.24	653	1.24
651	46	0.09	699	1.32
652	38	0.07	737	1.40
653	64	0.12	801	1.52
654	70	0.13	871	1.65
655	48	0.09	919	1.74
656	82	0.16	1001	1.90
657	95	0.18	1096	2.08
658	80	0.15	1176	2.23
659	66	0.12	1242	2.35
660	77	0.15	1319	2.50
661	79	0.15	1398	2.65
662	88	0.17	1486	2.81
663	85	0.16	1571	2.98
664	102	0.19	1673	3.17
665	152	0.29	1825	3.46
666	122	0.23	1947	3.69
667	146	0.28	2093	3.96
668	133	0.25	2226	4.22
669	150	0.28	2376	4.50
670	184	0.35	2560	4.85
671	196	0.37	2756	5.22
672	160	0.30	2916	5.52
673	197	0.37	3113	5.90
674	232	0.44	3345	6.33
675	214	0.41	3559	6.74
676	227	0.43	3786	7.17
677	288	0.55	4074	7.72
678	254	0.48	4328	8.20
679	232	0.44	4560	8.64
680	312	0.59	4872	9.23
681	308	0.58	5180	9.81
682	293	0.55	5473	10.36
683	328	0.62	5801	10.99
684	356	0.67	6157	11.66
685	349	0.66	6506	12.32
686	316	0.60	6822	12.92
687	386	0.73	7208	13.65
688	406	0.77	7614	14.42
689	380	0.72	7994	15.14
690	377	0.71	8371	15.85
691	366	0.69	8737	16.55
692	423	0.80	9160	17.35
693	437	0.83	9597	18.17
694	526	1.00	10123	19.17
695	464	0.88	10587	20.05
696	476	0.90	11063	20.95
697	435	0.82	11498	21.77
698	454	0.86	11952	22.63
699	511	0.97	12463	23.60
700	488	0.92	12951	24.53
701	511	0.97	13462	25.49
702	534	1.01	13996	26.51
703	479	0.91	14475	27.41
704	495	0.94	14970	28.35
705	492	0.93	15462	29.28
706	505	0.96	15967	30.24
707	488	0.92	16455	31.16
708	528	1.00	16983	32.16
709	473	0.90	17456	33.06
710	505	0.96	17961	34.01
711	454	0.86	18415	34.87
712	482	0.91	18897	35.79
713	503	0.95	19400	36.74
714	474	0.90	19874	37.64
715	513	0.97	20387	38.61
716	492	0.93	20879	39.54
717	467	0.88	21346	40.42
718	443	0.84	21789	41.26
719	460	0.87	22249	42.14
720	495	0.94	22744	43.07
721	455	0.86	23199	43.93
722	476	0.90	23675	44.84
723	484	0.92	24159	45.75
724	477	0.90	24636	46.66
725	489	0.93	25125	47.58
726	493	0.93	25618	48.52
727	490	0.93	26108	49.44
728	470	0.89	26578	50.33
729	438	0.83	27016	51.16
730	447	0.85	27463	52.01
731	479	0.91	27942	52.92
732	464	0.88	28406	53.80
733	425	0.80	28831	54.60
734	445	0.84	29276	55.44
735	456	0.86	29732	56.31
736	458	0.87	30190	57.17
737	418	0.79	30608	57.97
738	421	0.80	31029	58.76
739	451	0.85	31480	59.62
740	468	0.89	31948	60.50
741	412	0.78	32360	61.28
742	453	0.86	32813	62.14
743	415	0.79	33228	62.93
744	414	0.78	33642	63.71
745	409	0.77	34051	64.49
746	391	0.74	34442	65.23
747	416	0.79	34858	66.01
748	426	0.81	35284	66.82
749	424	0.80	35708	67.62

SS	Freq.	%	Cum. Freq.	Cum. %
750	433	0.82	36141	68.44
751	425	0.80	36566	69.25
752	441	0.84	37007	70.08
753	380	0.72	37387	70.80
754	373	0.71	37760	71.51
755	396	0.75	38156	72.26
756	428	0.81	38584	73.07
757	364	0.69	38948	73.76
758	410	0.78	39358	74.54
759	354	0.67	39712	75.21
760	371	0.70	40083	75.91
761	403	0.76	40486	76.67
762	311	0.59	40797	77.26
763	320	0.61	41117	77.87
764	356	0.67	41473	78.54
765	360	0.68	41833	79.22
766	324	0.61	42157	79.84
767	355	0.67	42512	80.51
768	305	0.58	42817	81.09
769	329	0.62	43146	81.71
770	346	0.66	43492	82.36
771	297	0.56	43789	82.93
772	288	0.55	44077	83.47
773	326	0.62	44403	84.09
774	309	0.59	44712	84.68
775	295	0.56	45007	85.23
776	287	0.54	45294	85.78
777	275	0.52	45569	86.30
778	265	0.50	45834	86.80
779	284	0.54	46118	87.34
780	258	0.49	46376	87.83
781	236	0.45	46612	88.27
782	281	0.53	46893	88.81
783	271	0.51	47164	89.32
784	220	0.42	47384	89.74
785	243	0.46	47627	90.20
786	237	0.45	47864	90.64
787	207	0.39	48071	91.04
788	169	0.32	48240	91.36
789	213	0.40	48453	91.76
790	192	0.36	48645	92.12
791	181	0.34	48826	92.47
792	178	0.34	49004	92.80
793	170	0.32	49174	93.13
794	156	0.30	49330	93.42
795	174	0.33	49504	93.75
796	147	0.28	49651	94.03
797	145	0.27	49796	94.30
798	128	0.24	49924	94.55
799	132	0.25	50056	94.80
800	131	0.25	50187	95.04
801	133	0.25	50320	95.30
802	117	0.22	50437	95.52
803	111	0.21	50548	95.73
804	89	0.17	50637	95.90
805	107	0.20	50744	96.10
806	112	0.21	50856	96.31
807	107	0.20	50963	96.51
808	76	0.14	51039	96.66
809	88	0.17	51127	96.82
810	65	0.12	51192	96.95
811	85	0.16	51277	97.11
812	73	0.14	51350	97.25
813	65	0.12	51415	97.37
814	60	0.11	51475	97.48
815	78	0.15	51553	97.63
816	64	0.12	51617	97.75
817	61	0.12	51678	97.87
818	52	0.10	51730	97.97
819	55	0.10	51785	98.07
820	51	0.10	51836	98.17
821	38	0.07	51874	98.24
822	56	0.11	51930	98.34
823	48	0.09	51978	98.44
824	54	0.10	52032	98.54
825	41	0.08	52073	98.62
826	40	0.08	52113	98.69
827	35	0.07	52148	98.76
828	43	0.08	52191	98.84
829	44	0.08	52235	98.92
830	33	0.06	52268	98.98
831	29	0.05	52297	99.04
832	24	0.05	52321	99.09
833	22	0.04	52343	99.13
834	23	0.04	52366	99.17
835	19	0.04	52385	99.21
836	28	0.05	52413	99.26
837	20	0.04	52433	99.30
838	18	0.03	52451	99.33
839	13	0.02	52464	99.36
840	12	0.02	52476	99.38
841	13	0.02	52489	99.40
842	11	0.02	52500	99.42
843	22	0.04	52522	99.47
844	14	0.03	52536	99.49
845	17	0.03	52553	99.52
846	21	0.04	52574	99.56
847	6	0.01	52580	99.58
848	19	0.04	52599	99.61
849	10	0.02	52609	99.63
850	195	0.37	52804	100.00

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	1000	1.82	1000	1.82
651	68	0.12	1068	1.94
652	80	0.15	1148	2.08
653	90	0.16	1238	2.25
654	113	0.21	1351	2.45
655	85	0.15	1436	2.61
656	110	0.20	1546	2.81
657	139	0.25	1685	3.06
658	124	0.23	1809	3.28
659	124	0.23	1933	3.51
660	129	0.23	2062	3.74
661	151	0.27	2213	4.02
662	159	0.29	2372	4.31
663	170	0.31	2542	4.62
664	178	0.32	2720	4.94
665	190	0.34	2910	5.28
666	189	0.34	3099	5.63
667	177	0.32	3276	5.95
668	205	0.37	3481	6.32
669	220	0.40	3701	6.72
670	220	0.40	3921	7.12
671	202	0.37	4123	7.49
672	258	0.47	4381	7.95
673	247	0.45	4628	8.40
674	279	0.51	4907	8.91
675	231	0.42	5138	9.33
676	261	0.47	5399	9.80
677	258	0.47	5657	10.27
678	256	0.46	5913	10.74
679	281	0.51	6194	11.25
680	273	0.50	6467	11.74
681	282	0.51	6749	12.25
682	297	0.54	7046	12.79
683	290	0.53	7336	13.32
684	261	0.47	7597	13.79
685	275	0.50	7872	14.29
686	297	0.54	8169	14.83
687	313	0.57	8482	15.40
688	294	0.53	8776	15.93
689	310	0.56	9086	16.50
690	302	0.55	9388	17.04
691	330	0.60	9718	17.64
692	284	0.52	10002	18.16
693	324	0.59	10326	18.75
694	296	0.54	10622	19.28
695	307	0.56	10929	19.84
696	310	0.56	11239	20.40
697	336	0.61	11575	21.01
698	299	0.54	11874	21.56
699	322	0.58	12196	22.14
700	328	0.60	12524	22.74
701	309	0.56	12833	23.30
702	326	0.59	13159	23.89
703	312	0.57	13471	24.46
704	350	0.64	13821	25.09
705	338	0.61	14159	25.71
706	355	0.64	14514	26.35
707	311	0.56	14825	26.91
708	341	0.62	15166	27.53
709	329	0.60	15495	28.13
710	406	0.74	15901	28.87
711	323	0.59	16224	29.45
712	374	0.68	16598	30.13
713	353	0.64	16951	30.77
714	364	0.66	17315	31.44
715	376	0.68	17691	32.12
716	367	0.67	18058	32.78
717	399	0.72	18457	33.51
718	394	0.72	18851	34.22
719	396	0.72	19247	34.94
720	393	0.71	19640	35.66
721	426	0.77	20066	36.43
722	388	0.70	20454	37.13
723	444	0.81	20898	37.94
724	422	0.77	21320	38.71
725	404	0.73	21724	39.44
726	436	0.79	22160	40.23
727	405	0.74	22565	40.97
728	432	0.78	22997	41.75
729	430	0.78	23427	42.53
730	423	0.77	23850	43.30
731	437	0.79	24287	44.09
732	465	0.84	24752	44.94
733	438	0.80	25190	45.73
734	452	0.82	25642	46.55
735	449	0.82	26091	47.37
736	460	0.84	26551	48.20
737	458	0.83	27009	49.04
738	503	0.91	27512	49.95
739	441	0.80	27953	50.75
740	430	0.78	28383	51.53
741	511	0.93	28894	52.46
742	417	0.76	29311	53.21
743	436	0.79	29747	54.01
744	506	0.92	30253	54.92
745	511	0.93	30764	55.85
746	442	0.80	31206	56.65
747	472	0.86	31678	57.51
748	491	0.89	32169	58.40
749	472	0.86	32641	59.26

SS	Freq.	%	Cum. Freq.	Cum. %
750	493	0.90	33134	60.16
751	466	0.85	33600	61.00
752	448	0.81	34048	61.81
753	469	0.85	34517	62.67
754	482	0.88	34999	63.54
755	478	0.87	35477	64.41
756	439	0.80	35916	65.21
757	453	0.82	36369	66.03
758	460	0.84	36829	66.86
759	457	0.83	37286	67.69
760	485	0.88	37771	68.57
761	437	0.79	38208	69.37
762	404	0.73	38612	70.10
763	456	0.83	39068	70.93
764	438	0.80	39506	71.72
765	466	0.85	39972	72.57
766	454	0.82	40426	73.39
767	439	0.80	40865	74.19
768	439	0.80	41304	74.99
769	411	0.75	41715	75.73
770	433	0.79	42148	76.52
771	367	0.67	42515	77.19
772	371	0.67	42886	77.86
773	383	0.70	43269	78.56
774	383	0.70	43652	79.25
775	345	0.63	43997	79.88
776	378	0.69	44375	80.56
777	346	0.63	44721	81.19
778	342	0.62	45063	81.81
779	347	0.63	45410	82.44
780	336	0.61	45746	83.05
781	334	0.61	46080	83.66
782	304	0.55	46384	84.21
783	328	0.60	46712	84.81
784	313	0.57	47025	85.37
785	287	0.52	47312	85.90
786	306	0.56	47618	86.45
787	266	0.48	47884	86.93
788	255	0.46	48139	87.40
789	285	0.52	48424	87.91
790	267	0.48	48691	88.40
791	260	0.47	48951	88.87
792	244	0.44	49195	89.31
793	233	0.42	49428	89.74
794	228	0.41	49656	90.15
795	214	0.39	49870	90.54
796	261	0.47	50131	91.01
797	217	0.39	50348	91.41
798	200	0.36	50548	91.77
799	183	0.33	50731	92.10
800	182	0.33	50913	92.43
801	199	0.36	51112	92.79
802	161	0.29	51273	93.09
803	194	0.35	51467	93.44
804	172	0.31	51639	93.75
805	154	0.28	51793	94.03
806	148	0.27	51941	94.30
807	155	0.28	52096	94.58
808	150	0.27	52246	94.85
809	148	0.27	52394	95.12
810	134	0.24	52528	95.37
811	117	0.21	52645	95.58
812	117	0.21	52762	95.79
813	127	0.23	52889	96.02
814	113	0.21	53002	96.23
815	114	0.21	53116	96.43
816	93	0.17	53209	96.60
817	105	0.19	53314	96.79
818	93	0.17	53407	96.96
819	86	0.16	53493	97.12
820	73	0.13	53566	97.25
821	97	0.18	53663	97.43
822	76	0.14	53739	97.56
823	74	0.13	53813	97.70
824	64	0.12	53877	97.81
825	58	0.11	53935	97.92
826	57	0.10	53992	98.02
827	50	0.09	54042	98.11
828	67	0.12	54109	98.24
829	50	0.09	54159	98.33
830	52	0.09	54211	98.42
831	57	0.10	54268	98.52
832	48	0.09	54316	98.61
833	42	0.08	54358	98.69
834	46	0.08	54404	98.77
835	42	0.08	54446	98.85
836	33	0.06	54479	98.91
837	38	0.07	54517	98.98
838	34	0.06	54551	99.04
839	35	0.06	54586	99.10
840	31	0.06	54617	99.16
841	26	0.05	54643	99.20
842	31	0.06	54674	99.26
843	21	0.04	54695	99.30
844	23	0.04	54718	99.34
845	23	0.04	54741	99.38
846	14	0.03	54755	99.41
847	19	0.03	54774	99.44
848	15	0.03	54789	99.47
849	11	0.02	54800	99.49
850	281	0.51	55081	100.00

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	496	0.89	496	0.89
651	45	0.08	541	0.97
652	45	0.08	586	1.05
653	28	0.05	614	1.10
654	44	0.08	658	1.18
655	59	0.11	717	1.29
656	62	0.11	779	1.40
657	80	0.14	859	1.54
658	55	0.10	914	1.64
659	69	0.12	983	1.76
660	82	0.15	1065	1.91
661	67	0.12	1132	2.03
662	80	0.14	1212	2.17
663	93	0.17	1305	2.34
664	104	0.19	1409	2.53
665	96	0.17	1505	2.70
666	95	0.17	1600	2.87
667	132	0.24	1732	3.11
668	143	0.26	1875	3.36
669	119	0.21	1994	3.58
670	137	0.25	2131	3.82
671	127	0.23	2258	4.05
672	146	0.26	2404	4.31
673	157	0.28	2561	4.59
674	154	0.28	2715	4.87
675	184	0.33	2899	5.20
676	172	0.31	3071	5.51
677	180	0.32	3251	5.83
678	176	0.32	3427	6.15
679	234	0.42	3661	6.57
680	182	0.33	3843	6.89
681	227	0.41	4070	7.30
682	233	0.42	4303	7.72
683	205	0.37	4508	8.09
684	223	0.40	4731	8.49
685	229	0.41	4960	8.90
686	225	0.40	5185	9.30
687	230	0.41	5415	9.71
688	260	0.47	5675	10.18
689	238	0.43	5913	10.61
690	229	0.41	6142	11.02
691	269	0.48	6411	11.50
692	267	0.48	6678	11.98
693	248	0.44	6926	12.42
694	252	0.45	7178	12.88
695	254	0.46	7432	13.33
696	235	0.42	7667	13.75
697	219	0.39	7886	14.15
698	253	0.45	8139	14.60
699	287	0.51	8426	15.11
700	274	0.49	8700	15.61
701	265	0.48	8965	16.08
702	271	0.49	9236	16.57
703	311	0.56	9547	17.13
704	289	0.52	9836	17.64
705	284	0.51	10120	18.15
706	309	0.55	10429	18.71
707	350	0.63	10779	19.34
708	308	0.55	11087	19.89
709	316	0.57	11403	20.46
710	300	0.54	11703	20.99
711	348	0.62	12051	21.62
712	337	0.60	12388	22.22
713	342	0.61	12730	22.84
714	367	0.66	13097	23.49
715	356	0.64	13453	24.13
716	366	0.66	13819	24.79
717	390	0.70	14209	25.49
718	388	0.70	14597	26.18
719	416	0.75	15013	26.93
720	378	0.68	15391	27.61
721	399	0.72	15790	28.32
722	430	0.77	16220	29.10
723	453	0.81	16673	29.91
724	442	0.79	17115	30.70
725	451	0.81	17566	31.51
726	481	0.86	18047	32.37
727	459	0.82	18506	33.20
728	518	0.93	19024	34.13
729	488	0.88	19512	35.00
730	496	0.89	20008	35.89
731	536	0.96	20544	36.85
732	536	0.96	21080	37.81
733	542	0.97	21622	38.79
734	543	0.97	22165	39.76
735	536	0.96	22701	40.72
736	529	0.95	23230	41.67
737	585	1.05	23815	42.72
738	579	1.04	24394	43.76
739	618	1.11	25012	44.87
740	539	0.97	25551	45.83
741	590	1.06	26141	46.89
742	598	1.07	26739	47.97
743	563	1.01	27302	48.98
744	586	1.05	27888	50.03
745	627	1.12	28515	51.15
746	644	1.16	29159	52.31
747	633	1.14	29792	53.44
748	660	1.18	30452	54.63
749	702	1.26	31154	55.89

SS	Freq.	%	Cum. Freq.	Cum. %
750	642	1.15	31796	57.04
751	666	1.19	32462	58.23
752	659	1.18	33121	59.41
753	636	1.14	33757	60.56
754	674	1.21	34431	61.76
755	676	1.21	35107	62.98
756	611	1.10	35718	64.07
757	690	1.24	36408	65.31
758	706	1.27	37114	66.58
759	608	1.09	37722	67.67
760	658	1.18	38380	68.85
761	642	1.15	39022	70.00
762	651	1.17	39673	71.17
763	679	1.22	40352	72.39
764	636	1.14	40988	73.53
765	635	1.14	41623	74.67
766	549	0.98	42172	75.65
767	575	1.03	42747	76.68
768	560	1.00	43307	77.69
769	576	1.03	43883	78.72
770	552	0.99	44435	79.71
771	534	0.96	44969	80.67
772	502	0.90	45471	81.57
773	490	0.88	45961	82.45
774	459	0.82	46420	83.27
775	455	0.82	46875	84.09
776	458	0.82	47333	84.91
777	430	0.77	47763	85.68
778	405	0.73	48168	86.41
779	416	0.75	48584	87.15
780	407	0.73	48991	87.88
781	379	0.68	49370	88.56
782	330	0.59	49700	89.15
783	343	0.62	50043	89.77
784	337	0.60	50380	90.37
785	294	0.53	50674	90.90
786	275	0.49	50949	91.39
787	266	0.48	51215	91.87
788	239	0.43	51454	92.30
789	240	0.43	51694	92.73
790	248	0.44	51942	93.18
791	223	0.40	52165	93.58
792	213	0.38	52378	93.96
793	195	0.35	52573	94.31
794	208	0.37	52781	94.68
795	205	0.37	52986	95.05
796	170	0.30	53156	95.35
797	158	0.28	53314	95.64
798	154	0.28	53468	95.91
799	142	0.25	53610	96.17
800	120	0.22	53730	96.38
801	140	0.25	53870	96.63
802	121	0.22	53991	96.85
803	139	0.25	54130	97.10
804	101	0.18	54231	97.28
805	103	0.18	54334	97.47
806	88	0.16	54422	97.62
807	88	0.16	54510	97.78
808	98	0.18	54608	97.96
809	87	0.16	54695	98.11
810	82	0.15	54777	98.26
811	62	0.11	54839	98.37
812	62	0.11	54901	98.48
813	55	0.10	54956	98.58
814	59	0.11	55015	98.69
815	51	0.09	55066	98.78
816	47	0.08	55113	98.86
817	53	0.10	55166	98.96
818	35	0.06	55201	99.02
819	32	0.06	55233	99.08
820	40	0.07	55273	99.15
821	41	0.07	55314	99.23
822	36	0.06	55350	99.29
823	28	0.05	55378	99.34
824	25	0.04	55403	99.38
825	22	0.04	55425	99.42
826	17	0.03	55442	99.45
827	21	0.04	55463	99.49
828	24	0.04	55487	99.54
829	25	0.04	55512	99.58
830	20	0.04	55532	99.62
831	13	0.02	55545	99.64
832	13	0.02	55558	99.66
833	18	0.03	55576	99.70
834	11	0.02	55587	99.71
835	15	0.03	55602	99.74
836	12	0.02	55614	99.76
837	16	0.03	55630	99.79
838	8	0.01	55638	99.81
839	17	0.03	55655	99.84
840	6	0.01	55661	99.85
841	4	0.01	55665	99.85
842	7	0.01	55672	99.87
843	6	0.01	55678	99.88
844	8	0.01	55686	99.89
845	4	0.01	55690	99.90
846	3	0.01	55693	99.90
847	8	0.01	55701	99.92
848	6	0.01	55707	99.93
849	3	0.01	55710	99.94
850	36	0.06	55746	100.00

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	153	0.27	153	0.27
651	13	0.02	166	0.29
652	5	0.01	171	0.30
653	21	0.04	192	0.33
654	10	0.02	202	0.35
655	17	0.03	219	0.38
656	14	0.02	233	0.41
657	7	0.01	240	0.42
658	12	0.02	252	0.44
659	17	0.03	269	0.47
660	14	0.02	283	0.49
661	14	0.02	297	0.52
662	25	0.04	322	0.56
663	23	0.04	345	0.60
664	25	0.04	370	0.65
665	35	0.06	405	0.71
666	27	0.05	432	0.75
667	26	0.05	458	0.80
668	34	0.06	492	0.86
669	26	0.05	518	0.90
670	38	0.07	556	0.97
671	45	0.08	601	1.05
672	45	0.08	646	1.13
673	53	0.09	699	1.22
674	47	0.08	746	1.30
675	63	0.11	809	1.41
676	64	0.11	873	1.52
677	64	0.11	937	1.63
678	74	0.13	1011	1.76
679	85	0.15	1096	1.91
680	88	0.15	1184	2.06
681	93	0.16	1277	2.23
682	94	0.16	1371	2.39
683	107	0.19	1478	2.58
684	115	0.20	1593	2.78
685	111	0.19	1704	2.97
686	108	0.19	1812	3.16
687	159	0.28	1971	3.44
688	159	0.28	2130	3.71
689	163	0.28	2293	4.00
690	178	0.31	2471	4.31
691	180	0.31	2651	4.62
692	191	0.33	2842	4.95
693	195	0.34	3037	5.29
694	260	0.45	3297	5.75
695	262	0.46	3559	6.20
696	274	0.48	3833	6.68
697	303	0.53	4136	7.21
698	292	0.51	4428	7.72
699	279	0.49	4707	8.21
700	309	0.54	5016	8.74
701	291	0.51	5307	9.25
702	331	0.58	5638	9.83
703	360	0.63	5998	10.46
704	351	0.61	6349	11.07
705	400	0.70	6749	11.77
706	388	0.68	7137	12.44
707	385	0.67	7522	13.11
708	429	0.75	7951	13.86
709	407	0.71	8358	14.57
710	409	0.71	8767	15.28
711	478	0.83	9245	16.12
712	443	0.77	9688	16.89
713	503	0.88	10191	17.77
714	471	0.82	10662	18.59
715	495	0.86	11157	19.45
716	510	0.89	11667	20.34
717	458	0.80	12125	21.14
718	514	0.90	12639	22.03
719	475	0.83	13114	22.86
720	517	0.90	13631	23.76
721	529	0.92	14160	24.69
722	536	0.93	14696	25.62
723	526	0.92	15222	26.54
724	577	1.01	15799	27.54
725	523	0.91	16322	28.45
726	573	1.00	16895	29.45
727	576	1.00	17471	30.46
728	615	1.07	18086	31.53
729	604	1.05	18690	32.58
730	600	1.05	19290	33.63
731	596	1.04	19886	34.67
732	638	1.11	20524	35.78
733	607	1.06	21131	36.84
734	571	1.00	21702	37.83
735	645	1.12	22347	38.96
736	586	1.02	22933	39.98
737	656	1.14	23589	41.12
738	620	1.08	24209	42.20
739	659	1.15	24868	43.35
740	643	1.12	25511	44.47
741	654	1.14	26165	45.61
742	632	1.10	26797	46.72
743	631	1.10	27428	47.82
744	657	1.15	28085	48.96
745	667	1.16	28752	50.12
746	668	1.16	29420	51.29
747	608	1.06	30028	52.35
748	636	1.11	30664	53.46
749	649	1.13	31313	54.59

SS	Freq.	%	Cum. Freq.	Cum. %
750	601	1.05	31914	55.64
751	645	1.12	32559	56.76
752	614	1.07	33173	57.83
753	620	1.08	33793	58.91
754	603	1.05	34396	59.96
755	646	1.13	35042	61.09
756	615	1.07	35657	62.16
757	615	1.07	36272	63.23
758	619	1.08	36891	64.31
759	654	1.14	37545	65.45
760	600	1.05	38145	66.50
761	568	0.99	38713	67.49
762	565	0.98	39278	68.47
763	628	1.09	39906	69.57
764	592	1.03	40498	70.60
765	505	0.88	41003	71.48
766	597	1.04	41600	72.52
767	595	1.04	42195	73.56
768	573	1.00	42768	74.56
769	569	0.99	43337	75.55
770	553	0.96	43890	76.51
771	551	0.96	44441	77.47
772	561	0.98	45002	78.45
773	495	0.86	45497	79.32
774	492	0.86	45989	80.17
775	492	0.86	46481	81.03
776	499	0.87	46980	81.90
777	523	0.91	47503	82.81
778	429	0.75	47932	83.56
779	449	0.78	48381	84.34
780	437	0.76	48818	85.11
781	457	0.80	49275	85.90
782	408	0.71	49683	86.61
783	404	0.70	50087	87.32
784	392	0.68	50479	88.00
785	408	0.71	50887	88.71
786	363	0.63	51250	89.34
787	363	0.63	51613	89.98
788	345	0.60	51958	90.58
789	310	0.54	52268	91.12
790	336	0.59	52604	91.71
791	298	0.52	52902	92.22
792	283	0.49	53185	92.72
793	280	0.49	53465	93.21
794	252	0.44	53717	93.65
795	244	0.43	53961	94.07
796	231	0.40	54192	94.47
797	226	0.39	54418	94.87
798	196	0.34	54614	95.21
799	188	0.33	54802	95.54
800	178	0.31	54980	95.85
801	174	0.30	55154	96.15
802	149	0.26	55303	96.41
803	152	0.26	55455	96.68
804	163	0.28	55618	96.96
805	124	0.22	55742	97.18
806	127	0.22	55869	97.40
807	123	0.21	55992	97.61
808	104	0.18	56096	97.79
809	87	0.15	56183	97.94
810	86	0.15	56269	98.09
811	80	0.14	56349	98.23
812	92	0.16	56441	98.39
813	89	0.16	56530	98.55
814	60	0.10	56590	98.65
815	60	0.10	56650	98.76
816	55	0.10	56705	98.85
817	51	0.09	56756	98.94
818	48	0.08	56804	99.03
819	47	0.08	56851	99.11
820	51	0.09	56902	99.20
821	33	0.06	56935	99.26
822	37	0.06	56972	99.32
823	29	0.05	57001	99.37
824	27	0.05	57028	99.42
825	24	0.04	57052	99.46
826	32	0.06	57084	99.52
827	28	0.05	57112	99.56
828	21	0.04	57133	99.60
829	15	0.03	57148	99.63
830	22	0.04	57170	99.67
831	19	0.03	57189	99.70
832	15	0.03	57204	99.72
833	20	0.03	57224	99.76
834	14	0.02	57238	99.78
835	13	0.02	57251	99.81
836	9	0.02	57260	99.82
837	7	0.01	57267	99.83
838	15	0.03	57282	99.86
839	5	0.01	57287	99.87
840	4	0.01	57291	99.88
841	8	0.01	57299	99.89
842	11	0.02	57310	99.91
843	9	0.02	57319	99.93
844	2	0.00	57321	99.93
845	6	0.01	57327	99.94
847	1	0.00	57328	99.94
848	1	0.00	57329	99.94
849	5	0.01	57334	99.95
850	28	0.05	57362	100.00

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	356	0.64	356	0.64
651	24	0.04	380	0.68
652	22	0.04	402	0.72
653	17	0.03	419	0.75
654	31	0.06	450	0.80
655	31	0.06	481	0.86
656	30	0.05	511	0.91
657	36	0.06	547	0.98
658	47	0.08	594	1.06
659	36	0.06	630	1.13
660	38	0.07	668	1.19
661	50	0.09	718	1.28
662	51	0.09	769	1.37
663	44	0.08	813	1.45
664	59	0.11	872	1.56
665	51	0.09	923	1.65
666	52	0.09	975	1.74
667	63	0.11	1038	1.85
668	78	0.14	1116	1.99
669	65	0.12	1181	2.11
670	106	0.19	1287	2.30
671	89	0.16	1376	2.46
672	85	0.15	1461	2.61
673	86	0.15	1547	2.76
674	103	0.18	1650	2.95
675	118	0.21	1768	3.16
676	117	0.21	1885	3.37
677	99	0.18	1984	3.55
678	134	0.24	2118	3.78
679	126	0.23	2244	4.01
680	135	0.24	2379	4.25
681	138	0.25	2517	4.50
682	163	0.29	2680	4.79
683	159	0.28	2839	5.07
684	152	0.27	2991	5.34
685	159	0.28	3150	5.63
686	191	0.34	3341	5.97
687	216	0.39	3557	6.36
688	210	0.38	3767	6.73
689	204	0.36	3971	7.10
690	200	0.36	4171	7.45
691	220	0.39	4391	7.85
692	233	0.42	4624	8.26
693	207	0.37	4831	8.63
694	260	0.46	5091	9.10
695	230	0.41	5321	9.51
696	272	0.49	5593	9.99
697	280	0.50	5873	10.49
698	265	0.47	6138	10.97
699	292	0.52	6430	11.49
700	281	0.50	6711	11.99
701	294	0.53	7005	12.52
702	320	0.57	7325	13.09
703	331	0.59	7656	13.68
704	332	0.59	7988	14.27
705	346	0.62	8334	14.89
706	362	0.65	8696	15.54
707	366	0.65	9062	16.19
708	392	0.70	9454	16.89
709	386	0.69	9840	17.58
710	403	0.72	10243	18.30
711	409	0.73	10652	19.04
712	427	0.76	11079	19.80
713	418	0.75	11497	20.55
714	410	0.73	11907	21.28
715	447	0.80	12354	22.08
716	485	0.87	12839	22.94
717	439	0.78	13278	23.73
718	497	0.89	13775	24.62
719	503	0.90	14278	25.51
720	485	0.87	14763	26.38
721	526	0.94	15289	27.32
722	497	0.89	15786	28.21
723	507	0.91	16293	29.12
724	498	0.89	16791	30.01
725	531	0.95	17322	30.95
726	548	0.98	17870	31.93
727	579	1.03	18449	32.97
728	524	0.94	18973	33.90
729	575	1.03	19548	34.93
730	560	1.00	20108	35.93
731	598	1.07	20706	37.00
732	604	1.08	21310	38.08
733	600	1.07	21910	39.15
734	614	1.10	22524	40.25
735	609	1.09	23133	41.34
736	585	1.05	23718	42.38
737	593	1.06	24311	43.44
738	581	1.04	24892	44.48
739	603	1.08	25495	45.56
740	644	1.15	26139	46.71
741	646	1.15	26785	47.86
742	636	1.14	27421	49.00
743	676	1.21	28097	50.21
744	628	1.12	28725	51.33
745	624	1.12	29349	52.45
746	660	1.18	30009	53.63
747	631	1.13	30640	54.75
748	616	1.10	31256	55.85
749	619	1.11	31875	56.96

SS	Freq.	%	Cum. Freq.	Cum. %
750	643	1.15	32518	58.11
751	665	1.19	33183	59.30
752	663	1.18	33846	60.48
753	570	1.02	34416	61.50
754	623	1.11	35039	62.61
755	633	1.13	35672	63.75
756	624	1.12	36296	64.86
757	638	1.14	36934	66.00
758	632	1.13	37566	67.13
759	628	1.12	38194	68.25
760	643	1.15	38837	69.40
761	604	1.08	39441	70.48
762	592	1.06	40033	71.54
763	590	1.05	40623	72.59
764	564	1.01	41187	73.60
765	540	0.96	41727	74.57
766	533	0.95	42260	75.52
767	597	1.07	42857	76.59
768	522	0.93	43379	77.52
769	500	0.89	43879	78.41
770	515	0.92	44394	79.33
771	460	0.82	44854	80.15
772	509	0.91	45363	81.06
773	493	0.88	45856	81.94
774	463	0.83	46319	82.77
775	478	0.85	46797	83.63
776	448	0.80	47245	84.43
777	444	0.79	47689	85.22
778	432	0.77	48121	85.99
779	410	0.73	48531	86.72
780	385	0.69	48916	87.41
781	376	0.67	49292	88.08
782	381	0.68	49673	88.77
783	367	0.66	50040	89.42
784	317	0.57	50357	89.99
785	306	0.55	50663	90.53
786	308	0.55	50971	91.08
787	293	0.52	51264	91.61
788	289	0.52	51553	92.12
789	265	0.47	51818	92.60
790	270	0.48	52088	93.08
791	247	0.44	52335	93.52
792	232	0.41	52567	93.94
793	211	0.38	52778	94.31
794	217	0.39	52995	94.70
795	190	0.34	53185	95.04
796	197	0.35	53382	95.39
797	193	0.34	53575	95.74
798	167	0.30	53742	96.04
799	151	0.27	53893	96.31
800	150	0.27	54043	96.57
801	123	0.22	54166	96.79
802	125	0.22	54291	97.02
803	125	0.22	54416	97.24
804	113	0.20	54529	97.44
805	97	0.17	54626	97.62
806	83	0.15	54709	97.76
807	84	0.15	54793	97.91
808	87	0.16	54880	98.07
809	94	0.17	54974	98.24
810	78	0.14	55052	98.38
811	79	0.14	55131	98.52
812	66	0.12	55197	98.64
813	66	0.12	55263	98.75
814	61	0.11	55324	98.86
815	58	0.10	55382	98.97
816	63	0.11	55445	99.08
817	40	0.07	55485	99.15
818	36	0.06	55521	99.22
819	40	0.07	55561	99.29
820	41	0.07	55602	99.36
821	25	0.04	55627	99.40
822	33	0.06	55660	99.46
823	22	0.04	55682	99.50
824	36	0.06	55718	99.57
825	30	0.05	55748	99.62
826	21	0.04	55769	99.66
827	23	0.04	55792	99.70
828	17	0.03	55809	99.73
829	16	0.03	55825	99.76
830	16	0.03	55841	99.79
831	5	0.01	55846	99.80
832	10	0.02	55856	99.81
833	11	0.02	55867	99.83
834	7	0.01	55874	99.85
835	9	0.02	55883	99.86
836	6	0.01	55889	99.87
837	4	0.01	55893	99.88
838	8	0.01	55901	99.89
839	5	0.01	55906	99.90
840	4	0.01	55910	99.91
841	6	0.01	55916	99.92
842	4	0.01	55920	99.93
843	2	0.00	55922	99.93
844	3	0.01	55925	99.94
845	3	0.01	55928	99.94
846	2	0.00	55930	99.95
847	3	0.01	55933	99.95
848	5	0.01	55938	99.96
850	22	0.04	55960	100.00

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	332	0.60	332	0.60
651	25	0.05	357	0.65
652	17	0.03	374	0.68
653	22	0.04	396	0.72
654	33	0.06	429	0.78
655	31	0.06	460	0.83
656	23	0.04	483	0.87
657	29	0.05	512	0.93
658	43	0.08	555	1.00
659	37	0.07	592	1.07
660	32	0.06	624	1.13
661	45	0.08	669	1.21
662	48	0.09	717	1.30
663	43	0.08	760	1.38
664	44	0.08	804	1.45
665	91	0.16	895	1.62
666	75	0.14	970	1.76
667	73	0.13	1043	1.89
668	104	0.19	1147	2.08
669	82	0.15	1229	2.22
670	98	0.18	1327	2.40
671	116	0.21	1443	2.61
672	111	0.20	1554	2.81
673	117	0.21	1671	3.02
674	140	0.25	1811	3.28
675	137	0.25	1948	3.52
676	145	0.26	2093	3.79
677	144	0.26	2237	4.05
678	158	0.29	2395	4.33
679	168	0.30	2563	4.64
680	207	0.37	2770	5.01
681	194	0.35	2964	5.36
682	200	0.36	3164	5.73
683	231	0.42	3395	6.14
684	226	0.41	3621	6.55
685	230	0.42	3851	6.97
686	252	0.46	4103	7.42
687	263	0.48	4366	7.90
688	267	0.48	4633	8.38
689	293	0.53	4926	8.91
690	313	0.57	5239	9.48
691	282	0.51	5521	9.99
692	353	0.64	5874	10.63
693	325	0.59	6199	11.22
694	325	0.59	6524	11.80
695	349	0.63	6873	12.44
696	331	0.60	7204	13.04
697	392	0.71	7596	13.74
698	376	0.68	7972	14.43
699	370	0.67	8342	15.09
700	397	0.72	8739	15.81
701	369	0.67	9108	16.48
702	366	0.66	9474	17.14
703	396	0.72	9870	17.86
704	359	0.65	10229	18.51
705	391	0.71	10620	19.22
706	385	0.70	11005	19.91
707	402	0.73	11407	20.64
708	416	0.75	11823	21.39
709	386	0.70	12209	22.09
710	380	0.69	12589	22.78
711	429	0.78	13018	23.56
712	438	0.79	13456	24.35
713	416	0.75	13872	25.10
714	440	0.80	14312	25.90
715	413	0.75	14725	26.64
716	457	0.83	15182	27.47
717	443	0.80	15625	28.27
718	453	0.82	16078	29.09
719	476	0.86	16554	29.95
720	457	0.83	17011	30.78
721	429	0.78	17440	31.56
722	460	0.83	17900	32.39
723	450	0.81	18350	33.20
724	456	0.83	18806	34.03
725	458	0.83	19264	34.86
726	487	0.88	19751	35.74
727	498	0.90	20249	36.64
728	500	0.90	20749	37.54
729	487	0.88	21236	38.43
730	520	0.94	21756	39.37
731	544	0.98	22300	40.35
732	519	0.94	22819	41.29
733	499	0.90	23318	42.19
734	515	0.93	23833	43.12
735	558	1.01	24391	44.13
736	488	0.88	24879	45.02
737	544	0.98	25423	46.00
738	518	0.94	25941	46.94
739	542	0.98	26483	47.92
740	562	1.02	27045	48.94
741	561	1.02	27606	49.95
742	574	1.04	28180	50.99
743	591	1.07	28771	52.06
744	564	1.02	29335	53.08
745	599	1.08	29934	54.16
746	575	1.04	30509	55.20
747	565	1.02	31074	56.23
748	546	0.99	31620	57.22
749	564	1.02	32184	58.24

SS	Freq.	%	Cum. Freq.	Cum. %
750	535	0.97	32719	59.20
751	558	1.01	33277	60.21
752	509	0.92	33786	61.13
753	518	0.94	34304	62.07
754	519	0.94	34823	63.01
755	552	1.00	35375	64.01
756	533	0.96	35908	64.97
757	542	0.98	36450	65.95
758	511	0.92	36961	66.88
759	541	0.98	37502	67.86
760	507	0.92	38009	68.78
761	505	0.91	38514	69.69
762	510	0.92	39024	70.61
763	524	0.95	39548	71.56
764	486	0.88	40034	72.44
765	493	0.89	40527	73.33
766	499	0.90	41026	74.24
767	474	0.86	41500	75.09
768	499	0.90	41999	76.00
769	447	0.81	42446	76.80
770	464	0.84	42910	77.64
771	438	0.79	43348	78.44
772	480	0.87	43828	79.31
773	432	0.78	44260	80.09
774	442	0.80	44702	80.89
775	425	0.77	45127	81.66
776	392	0.71	45519	82.36
777	440	0.80	45959	83.16
778	371	0.67	46330	83.83
779	407	0.74	46737	84.57
780	354	0.64	47091	85.21
781	338	0.61	47429	85.82
782	359	0.65	47788	86.47
783	362	0.66	48150	87.13
784	315	0.57	48465	87.70
785	351	0.64	48816	88.33
786	269	0.49	49085	88.82
787	265	0.48	49350	89.30
788	289	0.52	49639	89.82
789	297	0.54	49936	90.36
790	271	0.49	50207	90.85
791	268	0.48	50475	91.33
792	244	0.44	50719	91.77
793	235	0.43	50954	92.20
794	237	0.43	51191	92.63
795	215	0.39	51406	93.02
796	218	0.39	51624	93.41
797	193	0.35	51817	93.76
798	203	0.37	52020	94.13
799	183	0.33	52203	94.46
800	176	0.32	52379	94.78
801	196	0.35	52575	95.13
802	177	0.32	52752	95.45
803	167	0.30	52919	95.75
804	141	0.26	53060	96.01
805	140	0.25	53200	96.26
806	125	0.23	53325	96.49
807	135	0.24	53460	96.73
808	98	0.18	53558	96.91
809	111	0.20	53669	97.11
810	108	0.20	53777	97.31
811	89	0.16	53866	97.47
812	89	0.16	53955	97.63
813	83	0.15	54038	97.78
814	100	0.18	54138	97.96
815	90	0.16	54228	98.12
816	85	0.15	54313	98.28
817	63	0.11	54376	98.39
818	59	0.11	54435	98.50
819	62	0.11	54497	98.61
820	51	0.09	54548	98.70
821	54	0.10	54602	98.80
822	55	0.10	54657	98.90
823	48	0.09	54705	98.99
824	42	0.08	54747	99.06
825	44	0.08	54791	99.14
826	33	0.06	54824	99.20
827	28	0.05	54852	99.25
828	31	0.06	54883	99.31
829	34	0.06	54917	99.37
830	30	0.05	54947	99.42
831	19	0.03	54966	99.46
832	25	0.05	54991	99.50
833	22	0.04	55013	99.54
834	37	0.07	55050	99.61
835	16	0.03	55066	99.64
836	18	0.03	55084	99.67
837	18	0.03	55102	99.71
838	14	0.03	55116	99.73
839	12	0.02	55128	99.75
840	12	0.02	55140	99.77
841	10	0.02	55150	99.79
842	11	0.02	55161	99.81
843	6	0.01	55167	99.82
844	9	0.02	55176	99.84
845	10	0.02	55186	99.86
846	4	0.01	55190	99.86
847	5	0.01	55195	99.87
848	6	0.01	55201	99.88
849	7	0.01	55208	99.90
850	57	0.10	55265	100.00

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	436	0.83	436	0.83
651	37	0.07	473	0.90
652	34	0.06	507	0.96
653	48	0.09	555	1.05
654	40	0.08	595	1.13
655	60	0.11	655	1.24
656	38	0.07	693	1.31
657	64	0.12	757	1.44
658	46	0.09	803	1.52
659	63	0.12	866	1.64
660	65	0.12	931	1.77
661	62	0.12	993	1.88
662	63	0.12	1056	2.00
663	79	0.15	1135	2.15
664	101	0.19	1236	2.34
665	98	0.19	1334	2.53
666	108	0.20	1442	2.73
667	110	0.21	1552	2.94
668	97	0.18	1649	3.13
669	116	0.22	1765	3.35
670	126	0.24	1891	3.59
671	126	0.24	2017	3.83
672	148	0.28	2165	4.11
673	163	0.31	2328	4.42
674	149	0.28	2477	4.70
675	202	0.38	2679	5.08
676	171	0.32	2850	5.41
677	187	0.35	3037	5.76
678	230	0.44	3267	6.20
679	186	0.35	3453	6.55
680	224	0.42	3677	6.97
681	186	0.35	3863	7.33
682	218	0.41	4081	7.74
683	228	0.43	4309	8.17
684	227	0.43	4536	8.60
685	229	0.43	4765	9.04
686	270	0.51	5035	9.55
687	257	0.49	5292	10.04
688	243	0.46	5535	10.50
689	260	0.49	5795	10.99
690	252	0.48	6047	11.47
691	274	0.52	6321	11.99
692	268	0.51	6589	12.50
693	299	0.57	6888	13.06
694	301	0.57	7189	13.63
695	298	0.57	7487	14.20
696	301	0.57	7788	14.77
697	317	0.60	8105	15.37
698	315	0.60	8420	15.97
699	359	0.68	8779	16.65
700	302	0.57	9081	17.22
701	305	0.58	9386	17.80
702	329	0.62	9715	18.43
703	319	0.61	10034	19.03
704	373	0.71	10407	19.74
705	344	0.65	10751	20.39
706	340	0.64	11091	21.03
707	302	0.57	11393	21.61
708	302	0.57	11695	22.18
709	347	0.66	12042	22.84
710	331	0.63	12373	23.47
711	301	0.57	12674	24.04
712	337	0.64	13011	24.68
713	357	0.68	13368	25.35
714	354	0.67	13722	26.02
715	350	0.66	14072	26.69
716	381	0.72	14453	27.41
717	419	0.79	14872	28.21
718	415	0.79	15287	28.99
719	390	0.74	15677	29.73
720	398	0.75	16075	30.49
721	407	0.77	16482	31.26
722	415	0.79	16897	32.05
723	436	0.83	17333	32.87
724	451	0.86	17784	33.73
725	424	0.80	18208	34.53
726	413	0.78	18621	35.32
727	439	0.83	19060	36.15
728	455	0.86	19515	37.01
729	470	0.89	19985	37.90
730	442	0.84	20427	38.74
731	451	0.86	20878	39.60
732	457	0.87	21335	40.46
733	461	0.87	21796	41.34
734	467	0.89	22263	42.22
735	452	0.86	22715	43.08
736	454	0.86	23169	43.94
737	472	0.90	23641	44.84
738	494	0.94	24135	45.77
739	438	0.83	24573	46.60
740	469	0.89	25042	47.49
741	509	0.97	25551	48.46
742	509	0.97	26060	49.42
743	510	0.97	26570	50.39
744	528	1.00	27098	51.39
745	480	0.91	27578	52.30
746	487	0.92	28065	53.23
747	502	0.95	28567	54.18
748	510	0.97	29077	55.15
749	483	0.92	29560	56.06

SS	Freq.	%	Cum. Freq.	Cum. %
750	473	0.90	30033	56.96
751	476	0.90	30509	57.86
752	511	0.97	31020	58.83
753	463	0.88	31483	59.71
754	467	0.89	31950	60.60
755	494	0.94	32444	61.53
756	462	0.88	32906	62.41
757	463	0.88	33369	63.29
758	460	0.87	33829	64.16
759	473	0.90	34302	65.06
760	492	0.93	34794	65.99
761	471	0.89	35265	66.88
762	444	0.84	35709	67.72
763	447	0.85	36156	68.57
764	459	0.87	36615	69.44
765	441	0.84	37056	70.28
766	470	0.89	37526	71.17
767	434	0.82	37960	71.99
768	431	0.82	38391	72.81
769	437	0.83	38828	73.64
770	443	0.84	39271	74.48
771	435	0.83	39706	75.30
772	439	0.83	40145	76.14
773	415	0.79	40560	76.92
774	437	0.83	40997	77.75
775	419	0.79	41416	78.55
776	391	0.74	41807	79.29
777	353	0.67	42160	79.96
778	388	0.74	42548	80.69
779	390	0.74	42938	81.43
780	349	0.66	43287	82.10
781	330	0.63	43617	82.72
782	277	0.53	43894	83.25
783	346	0.66	44240	83.90
784	309	0.59	44549	84.49
785	339	0.64	44888	85.13
786	321	0.61	45209	85.74
787	306	0.58	45515	86.32
788	280	0.53	45795	86.85
789	254	0.48	46049	87.33
790	285	0.54	46334	87.88
791	266	0.50	46600	88.38
792	260	0.49	46860	88.87
793	251	0.48	47111	89.35
794	227	0.43	47338	89.78
795	226	0.43	47564	90.21
796	227	0.43	47791	90.64
797	235	0.45	48026	91.08
798	204	0.39	48230	91.47
799	185	0.35	48415	91.82
800	194	0.37	48609	92.19
801	191	0.36	48800	92.55
802	180	0.34	48980	92.89
803	166	0.31	49146	93.21
804	172	0.33	49318	93.53
805	164	0.31	49482	93.85
806	157	0.30	49639	94.14
807	154	0.29	49793	94.44
808	161	0.31	49954	94.74
809	118	0.22	50072	94.96
810	132	0.25	50204	95.21
811	127	0.24	50331	95.46
812	109	0.21	50440	95.66
813	92	0.17	50532	95.84
814	114	0.22	50646	96.05
815	95	0.18	50741	96.23
816	108	0.20	50849	96.44
817	93	0.18	50942	96.61
818	97	0.18	51039	96.80
819	89	0.17	51128	96.97
820	80	0.15	51208	97.12
821	81	0.15	51289	97.27
822	91	0.17	51380	97.45
823	92	0.17	51472	97.62
824	65	0.12	51537	97.74
825	61	0.12	51598	97.86
826	73	0.14	51671	98.00
827	66	0.13	51737	98.12
828	62	0.12	51799	98.24
829	57	0.11	51856	98.35
830	55	0.10	51911	98.45
831	65	0.12	51976	98.58
832	55	0.10	52031	98.68
833	40	0.08	52071	98.76
834	41	0.08	52112	98.83
835	48	0.09	52160	98.92
836	31	0.06	52191	98.98
837	46	0.09	52237	99.07
838	36	0.07	52273	99.14
839	38	0.07	52311	99.21
840	27	0.05	52338	99.26
841	29	0.06	52367	99.32
842	27	0.05	52394	99.37
843	29	0.06	52423	99.42
844	15	0.03	52438	99.45
845	20	0.04	52458	99.49
846	17	0.03	52475	99.52
847	23	0.04	52498	99.57
848	15	0.03	52513	99.59
849	19	0.04	52532	99.63
850	195	0.37	52727	100.00

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	12	0.92	12	0.92
665	13	1.00	25	1.92
676	31	2.38	56	4.30
684	39	3.00	95	7.30
690	42	3.23	137	10.53
695	52	4.00	189	14.53
699	51	3.92	240	18.45
700	56	4.30	296	22.75
706	40	3.07	336	25.83
709	51	3.92	387	29.75
712	38	2.92	425	32.67
714	55	4.23	480	36.89
717	39	3.00	519	39.89
719	32	2.46	551	42.35
721	28	2.15	579	44.50
723	25	1.92	604	46.43
725	37	2.84	641	49.27
726	30	2.31	671	51.58
728	29	2.23	700	53.80
730	21	1.61	721	55.42
731	25	1.92	746	57.34
733	40	3.07	786	60.42
734	23	1.77	809	62.18
736	29	2.23	838	64.41
737	31	2.38	869	66.79
739	26	2.00	895	68.79
740	19	1.46	914	70.25
742	16	1.23	930	71.48
743	30	2.31	960	73.79
745	22	1.69	982	75.48
746	17	1.31	999	76.79
747	21	1.61	1020	78.40
749	23	1.77	1043	80.17
750	22	1.69	1065	81.86
752	23	1.77	1088	83.63
753	19	1.46	1107	85.09
755	22	1.69	1129	86.78
756	21	1.61	1150	88.39
758	12	0.92	1162	89.32
760	15	1.15	1177	90.47
761	11	0.85	1188	91.31
763	9	0.69	1197	92.01
765	14	1.08	1211	93.08
766	14	1.08	1225	94.16
768	15	1.15	1240	95.31
770	14	1.08	1254	96.39
772	8	0.61	1262	97.00
774	4	0.31	1266	97.31
776	6	0.46	1272	97.77
779	3	0.23	1275	98.00
781	9	0.69	1284	98.69
784	3	0.23	1287	98.92
786	6	0.46	1293	99.39
793	5	0.38	1298	99.77
796	1	0.08	1299	99.85
800	1	0.08	1300	99.92
809	1	0.08	1301	100.00

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

SS	Freq.	%	Cum. Freq.	Cum. %
650	5	0.43	5	0.43
661	3	0.26	8	0.70
672	10	0.87	18	1.56
680	9	0.78	27	2.35
686	14	1.22	41	3.56
690	33	2.87	74	6.43
695	33	2.87	107	9.30
698	26	2.26	133	11.56
700	37	3.21	170	14.77
704	35	3.04	205	17.81
707	35	3.04	240	20.85
709	36	3.13	276	23.98
711	31	2.69	307	26.67
713	32	2.78	339	29.45
715	28	2.43	367	31.89
717	26	2.26	393	34.14
719	24	2.09	417	36.23
721	31	2.69	448	38.92
722	36	3.13	484	42.05
724	35	3.04	519	45.09
725	33	2.87	552	47.96
727	33	2.87	585	50.83
728	29	2.52	614	53.34
730	34	2.95	648	56.30
731	22	1.91	670	58.21
733	33	2.87	703	61.08
734	22	1.91	725	62.99
735	24	2.09	749	65.07
736	30	2.61	779	67.68
738	18	1.56	797	69.24
739	35	3.04	832	72.28
740	26	2.26	858	74.54
742	20	1.74	878	76.28
743	20	1.74	898	78.02
744	17	1.48	915	79.50
745	25	2.17	940	81.67
747	11	0.96	951	82.62
748	20	1.74	971	84.36
749	22	1.91	993	86.27
750	14	1.22	1007	87.49
752	15	1.30	1022	88.79
753	17	1.48	1039	90.27
754	10	0.87	1049	91.14
756	17	1.48	1066	92.62
757	11	0.96	1077	93.57
759	11	0.96	1088	94.53
760	7	0.61	1095	95.13
762	10	0.87	1105	96.00
763	7	0.61	1112	96.61
765	11	0.96	1123	97.57
766	4	0.35	1127	97.91
768	6	0.52	1133	98.44
770	4	0.35	1137	98.78
772	1	0.09	1138	98.87
774	4	0.35	1142	99.22
776	2	0.17	1144	99.39
778	2	0.17	1146	99.57
780	1	0.09	1147	99.65
782	1	0.09	1148	99.74
785	1	0.09	1149	99.83
788	2	0.17	1151	100.00

Appendix G: Scale Score Distribution Graphs

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

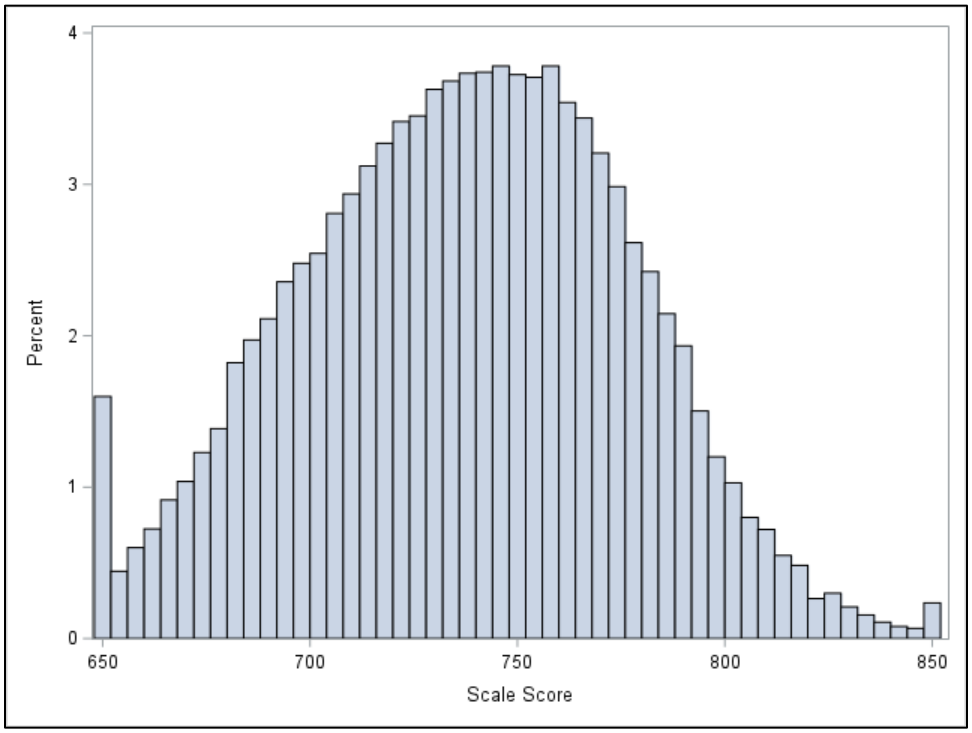


Figure G.2. Scale Score Distribution—Mathematics Grade 4

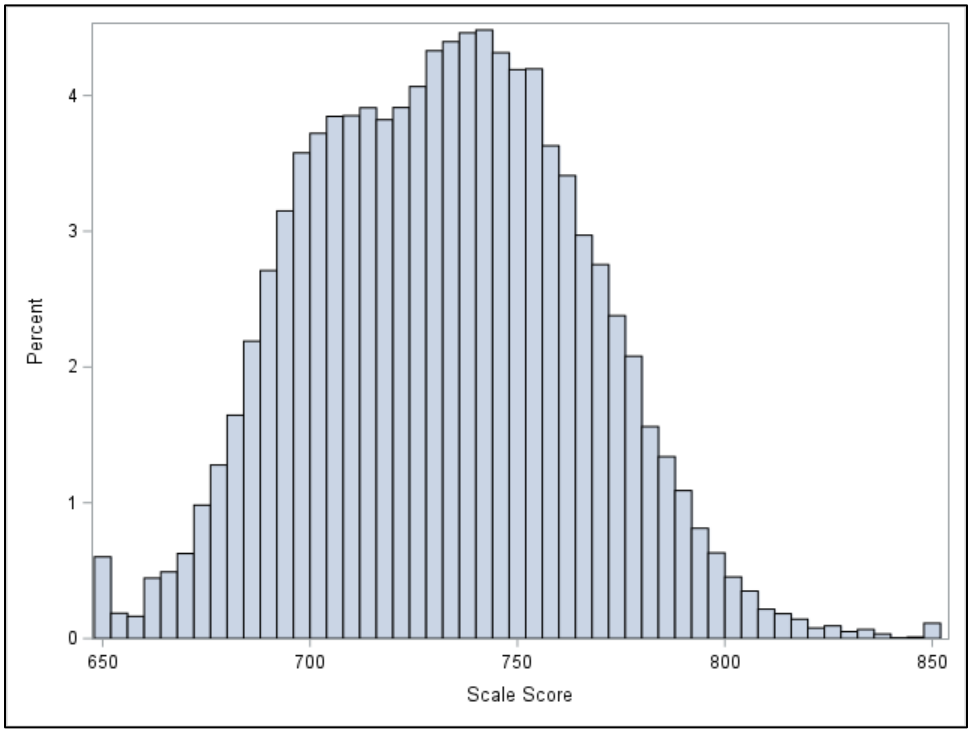


Figure G.3. Scale Score Distribution—Mathematics Grade 5

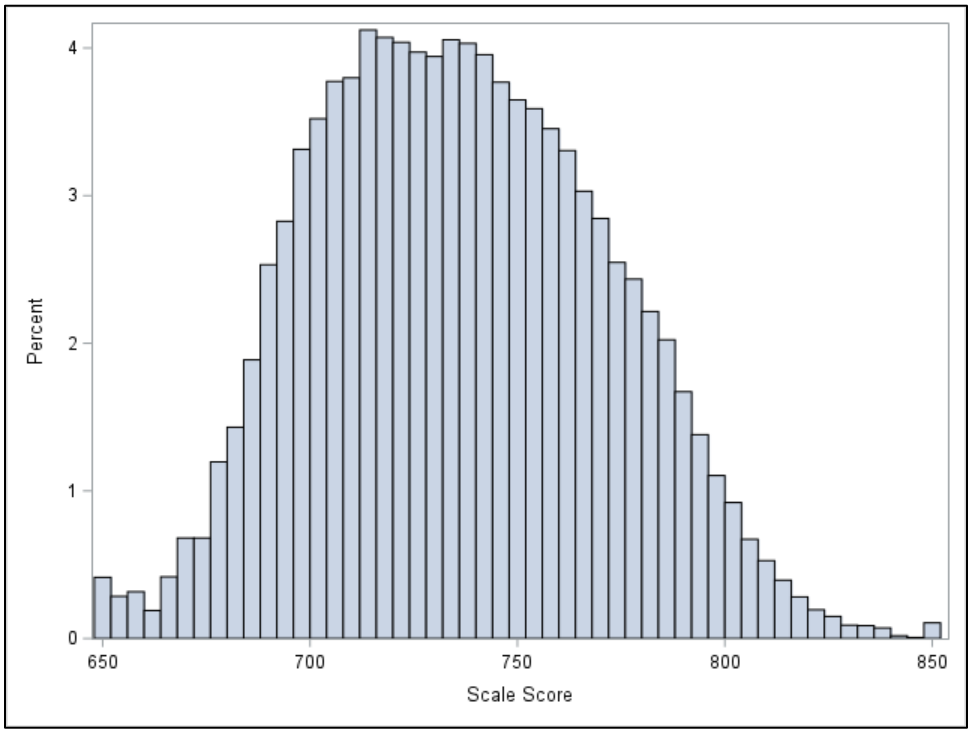


Figure G.4. Scale Score Distribution—Mathematics Grade 6

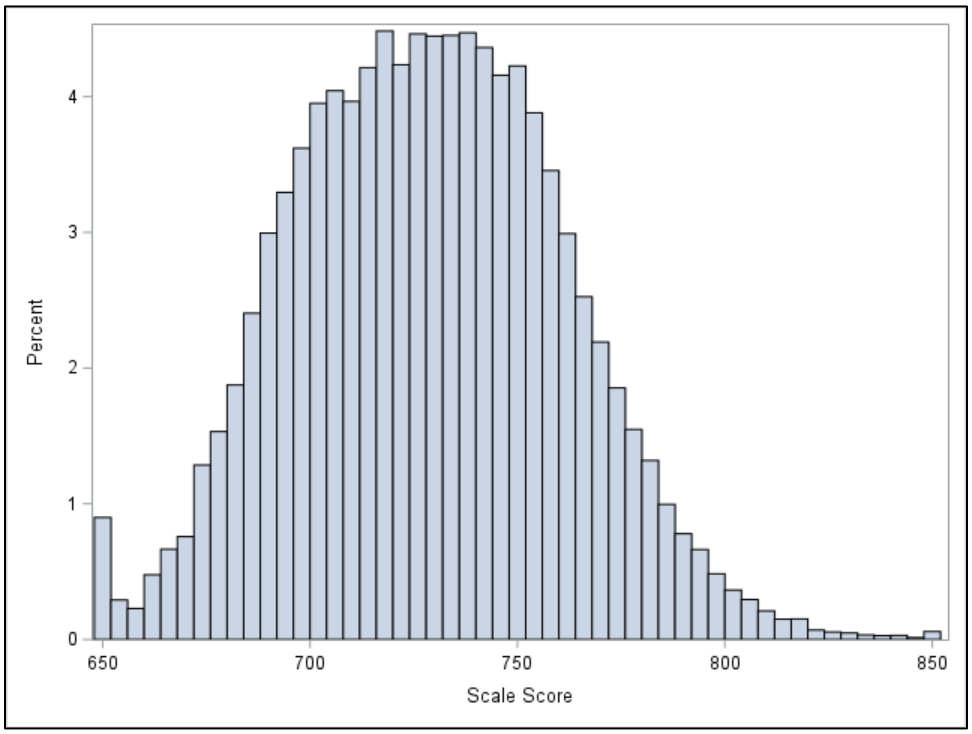


Figure G.5. Scale Score Distribution—Mathematics Grade 7

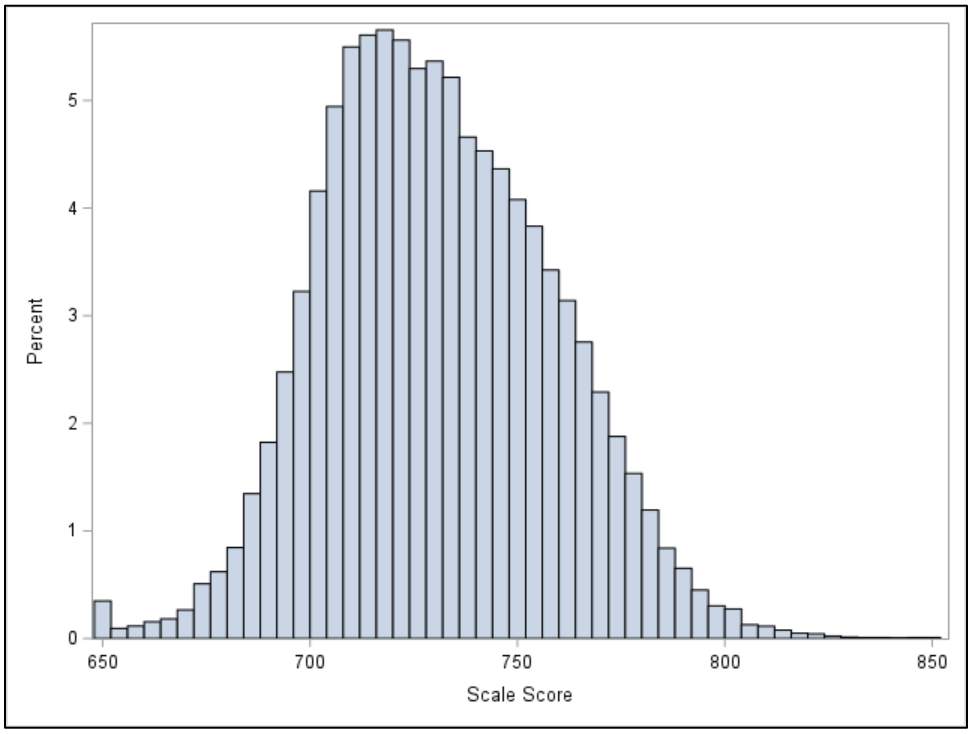


Figure G.6. Scale Score Distribution—Mathematics Grade 8

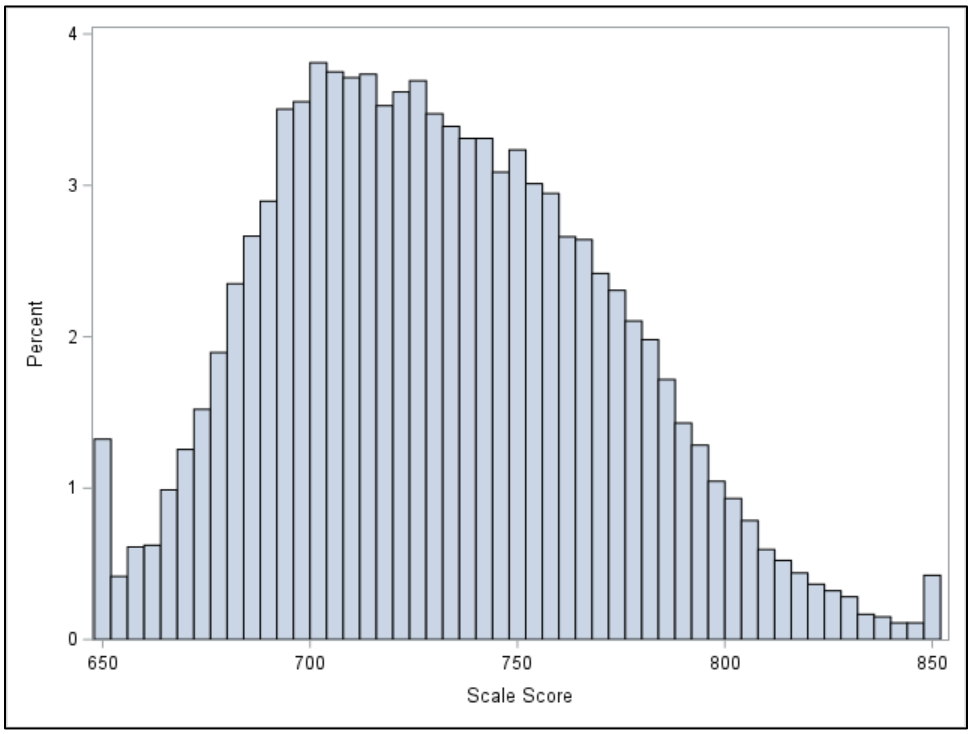


Figure G.7. Scale Score Distribution—ELA Grade 3

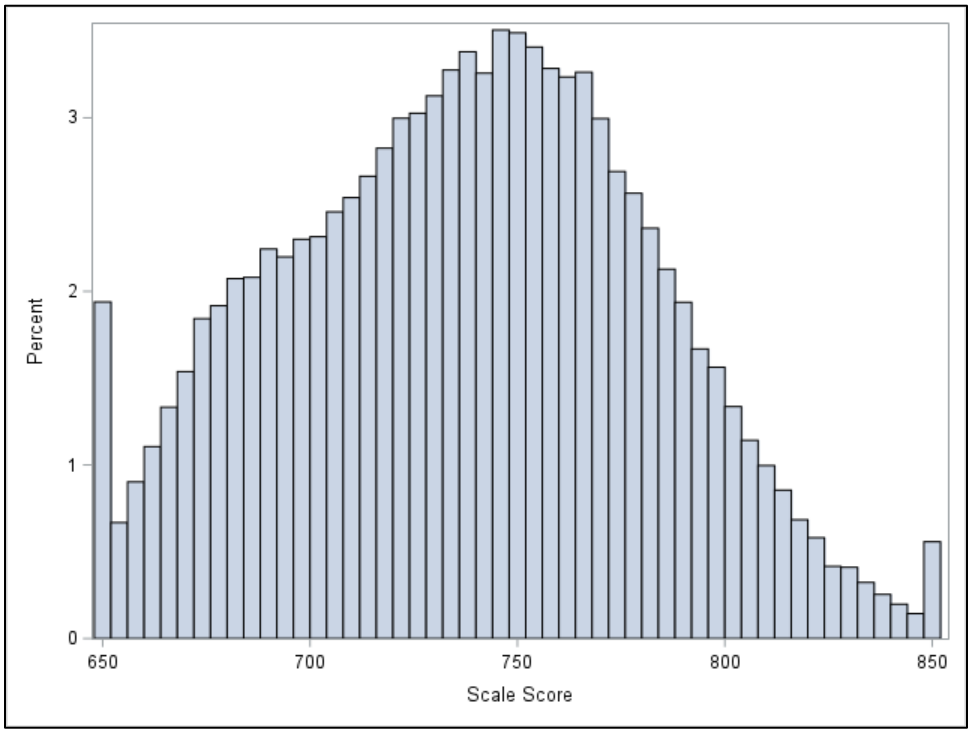


Figure G.8. Scale Score Distribution—ELA Grade 4

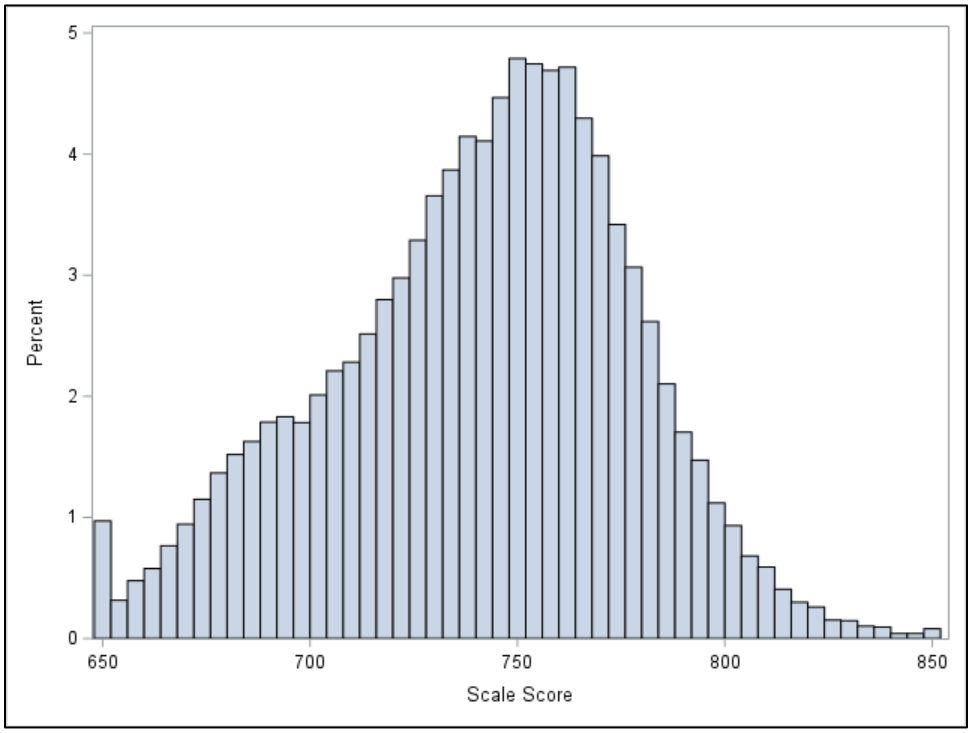


Figure G.9. Scale Score Distribution—ELA Grade 5

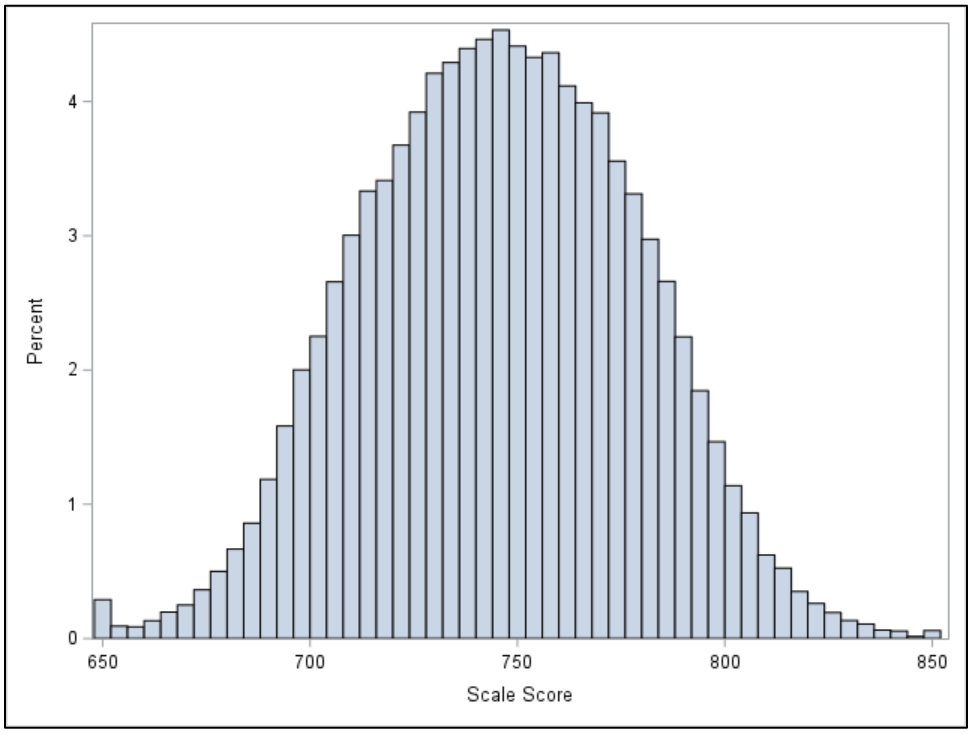


Figure G.10. Scale Score Distribution—ELA Grade 6

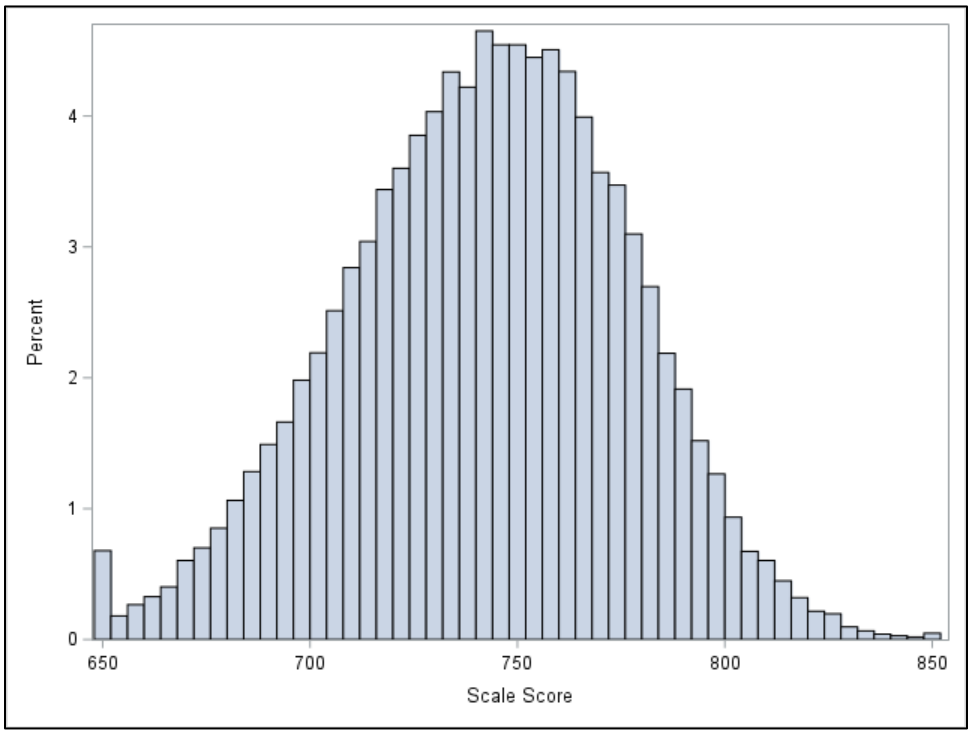


Figure G.11. Scale Score Distribution—ELA Grade 7

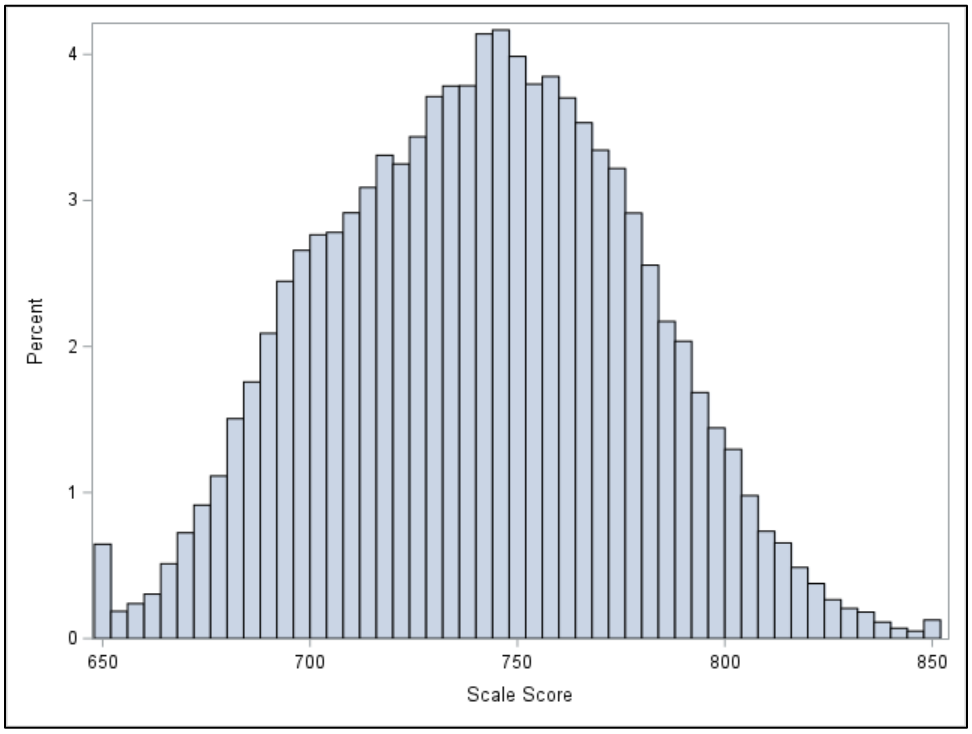


Figure G.12. Scale Score Distribution—ELA Grade 8

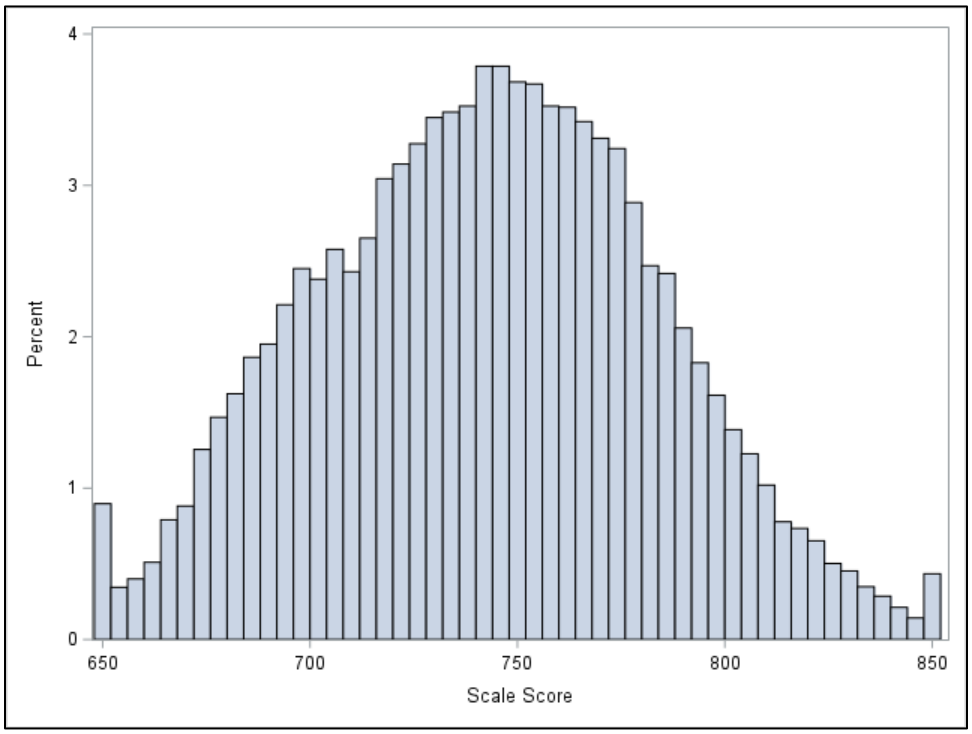


Figure G.13. Scale Score Distribution—CSLA Grade 3

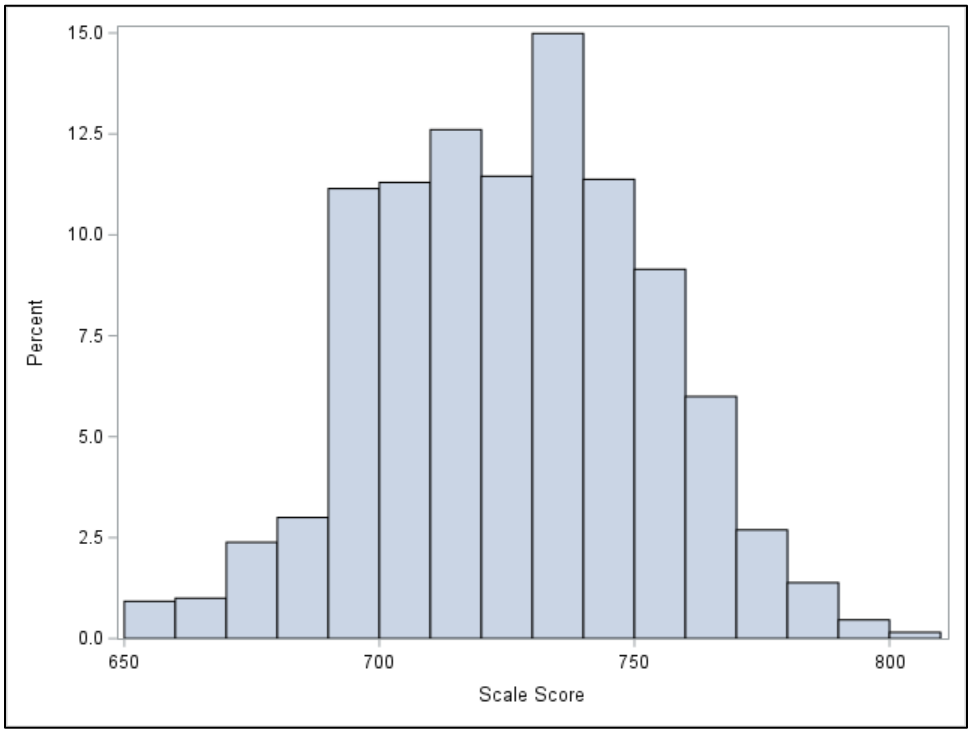
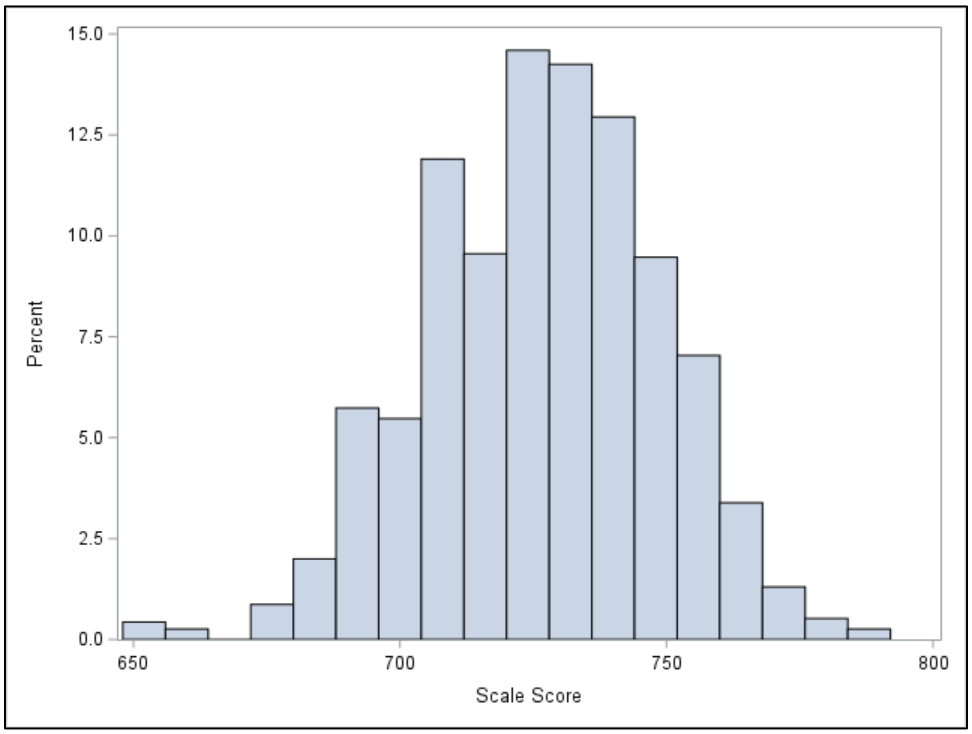


Figure G.14. Scale Score Distribution—CSLA Grade 4



Appendix H: Scale Score Summary Statistics by Demographic Group

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,833	741.06	37.77	650	850	0.91
IEP	6,649	706.06	37.07	650	850	0.93
No Accommodation	43,595	744.68	37.04	650	850	0.91
Accommodation	12,887	710.77	35.43	650	850	0.92
Am. Indian/Alaska Native	359	715.33	36.17	650	835	0.91
Asian	1,973	758.17	39.75	650	850	0.91
Black	2,572	719.68	38.12	650	843	0.91
Hispanic	18,976	719.31	36.39	650	850	0.91
White	29,415	748.13	36.15	650	850	0.90
Hawaiian/Pacific Islander	185	717.48	36.07	650	809	0.91
Two or More Races	2,995	743.42	39.50	650	850	0.91
Missing	*	*	*	*	*	*
No Economic Disadvantage	37,119	746.68	37.61	650	850	0.90
Economic Disadvantage	19,363	718.27	35.67	650	850	0.91
Female	27,925	735.30	38.48	650	850	0.91
Male	28,557	738.55	40.09	650	850	0.91
Language Proficiency NA	46,647	741.41	38.39	650	850	0.91
Language Proficiency NEP	2,619	691.48	29.70	650	850	0.87
Language Proficiency LEP	6,181	718.97	32.46	650	850	0.90
Language Proficiency FEP	1,035	758.03	31.71	650	850	0.87
Not Migrant	56,310	737.03	39.31	650	850	0.91
Migrant	172	708.73	37.65	650	797	0.92

*n-count less than 16

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	50,155	735.28	31.95	650	850	0.90
IEP	6,731	705.45	27.92	650	850	0.91
No Accommodation	44,581	737.97	31.49	650	850	0.90
Accommodation	12,305	709.21	27.84	650	850	0.91
Am. Indian/Alaska Native	364	715.39	31.57	650	850	0.91
Asian	1,898	749.07	33.72	650	850	0.91
Black	2,507	715.66	29.14	650	837	0.89
Hispanic	19,428	716.92	29.25	650	847	0.89
White	29,601	741.60	31.10	650	850	0.89
Hawaiian/Pacific Islander	164	718.37	31.75	650	801	0.91
Two or More Races	2,919	735.92	32.69	650	850	0.91
Missing	*	*	*	*	*	*
No Economic Disadvantage	37,276	740.01	32.01	650	850	0.90
Economic Disadvantage	19,610	716.04	28.69	650	850	0.89

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Female	27,698	729.49	31.92	650	850	0.90
Male	29,188	733.89	33.74	650	850	0.91
Language Proficiency NA	46,588	735.77	32.30	650	850	0.90
Language Proficiency NEP	2,005	694.14	21.59	650	790	0.84
Language Proficiency LEP	6,245	711.11	25.51	650	821	0.87
Language Proficiency FEP	2,048	739.98	29.07	650	850	0.89
Not Migrant	56,733	731.80	32.95	650	850	0.91
Migrant	153	711.39	23.13	651	777	0.85

*n-count less than 16

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	50,481	739.50	34.06	650	850	0.92
IEP	6,942	706.91	28.71	650	850	0.92
No Accommodation	45,866	741.50	33.98	650	850	0.92
Accommodation	11,557	711.99	29.11	650	850	0.92
Am. Indian/Alaska Native	347	717.54	32.20	650	850	0.92
Asian	1,957	755.93	36.00	650	850	0.92
Black	2,509	719.40	31.49	650	850	0.91
Hispanic	20,045	720.56	30.71	650	850	0.91
White	29,446	745.54	33.66	650	850	0.91
Hawaiian/Pacific Islander	181	720.64	32.96	650	819	0.92
Two or More Races	2,934	741.25	35.79	650	850	0.92
Missing	*	*	*	*	*	*
No Economic Disadvantage	37,866	743.84	34.57	650	850	0.91
Economic Disadvantage	19,557	719.52	30.25	650	850	0.91
Female	28,087	733.49	34.15	650	850	0.92
Male	29,336	737.54	35.89	650	850	0.92
Language Proficiency NA	46,647	739.42	34.97	650	850	0.92
Language Proficiency NEP	1,325	698.28	22.53	650	814	0.83
Language Proficiency LEP	5,756	711.38	25.66	650	817	0.88
Language Proficiency FEP	3,695	737.80	30.17	650	850	0.90
Not Migrant	57,252	735.62	35.11	650	850	0.92
Migrant	171	715.20	26.83	650	787	0.87

*n-count less than 16

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,513	731.80	31.41	650	850	0.90
IEP	6,423	699.43	26.23	650	824	0.88
No Accommodation	48,943	731.80	31.53	650	850	0.90
Accommodation	6,993	702.08	27.17	650	847	0.89
Am. Indian/Alaska Native	359	710.68	28.79	650	803	0.87
Asian	1,740	748.75	35.22	650	850	0.91
Black	2,489	711.80	29.77	650	832	0.88
Hispanic	19,998	714.00	28.37	650	850	0.87
White	28,466	737.96	30.74	650	850	0.89
Hawaiian/Pacific Islander	142	711.61	28.13	650	776	0.86
Two or More Races	2,736	733.09	33.51	650	850	0.91
Missing	*	*	*	*	*	*
No Economic Disadvantage	36,694	736.24	31.87	650	850	0.90
Economic Disadvantage	19,242	712.53	27.77	650	850	0.86
Female	27,334	726.45	31.92	650	850	0.89
Male	28,602	729.64	33.04	650	850	0.90
Language Proficiency NA	45,325	732.13	32.21	650	850	0.90
Language Proficiency NEP	1,111	691.43	21.79	650	777	0.77
Language Proficiency LEP	4,366	700.88	22.85	650	804	0.79
Language Proficiency FEP	5,134	723.40	27.19	650	850	0.87
Not Migrant	55,775	728.14	32.54	650	850	0.90
Migrant	161	709.48	24.61	650	776	0.83

*n-count less than 16

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,285	733.15	27.09	650	850	0.89
IEP	5,998	707.19	20.74	650	818	0.84
No Accommodation	48,669	733.10	27.17	650	850	0.89
Accommodation	6,614	710.00	22.41	650	824	0.88
Am. Indian/Alaska Native	366	717.40	24.41	651	805	0.87
Asian	1,766	748.94	30.37	650	850	0.91
Black	2,419	718.51	25.56	650	821	0.87
Hispanic	20,283	718.78	23.98	650	840	0.86
White	27,744	738.55	26.54	650	850	0.89
Hawaiian/Pacific Islander	159	717.68	24.16	656	783	0.87
Two or More Races	2,539	733.95	27.90	650	830	0.90
Missing	*	*	*	*	*	*
No Economic Disadvantage	36,591	736.80	27.42	650	850	0.90
Economic Disadvantage	18,692	717.68	23.50	650	839	0.85
Female	26,597	729.59	27.67	650	850	0.89
Male	28,686	731.03	27.67	650	850	0.90

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Language Proficiency NA	44,726	733.80	27.48	650	850	0.89
Language Proficiency NEP	1,219	700.80	18.54	650	771	0.64
Language Proficiency LEP	4,020	708.13	19.23	650	811	0.74
Language Proficiency FEP	5,318	724.81	23.48	650	831	0.86
Not Migrant	55,120	730.39	27.68	650	850	0.89
Migrant	163	713.64	21.08	660	785	0.79

*n-count less than 16

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	47,307	735.28	39.39	650	850	0.90
IEP	5,497	696.81	27.38	650	850	0.84
No Accommodation	46,893	734.91	39.55	650	850	0.90
Accommodation	5,911	702.41	31.60	650	850	0.88
Am. Indian/Alaska Native	312	711.08	36.18	650	825	0.89
Asian	1,705	758.76	43.51	650	850	0.91
Black	2,292	713.72	34.87	650	850	0.88
Hispanic	19,685	713.70	33.42	650	850	0.86
White	26,262	743.90	38.83	650	850	0.90
Hawaiian/Pacific Islander	155	713.09	32.35	650	804	0.86
Two or More Races	2,386	738.34	41.34	650	850	0.91
Missing	*	*	*	*	*	*
No Economic Disadvantage	34,963	740.71	40.00	650	850	0.90
Economic Disadvantage	17,841	712.77	33.17	650	850	0.86
Female	25,123	731.78	39.41	650	850	0.90
Male	27,681	730.82	40.66	650	850	0.91
Language Proficiency NA	43,528	736.58	39.88	650	850	0.90
Language Proficiency NEP	1,169	689.64	22.22	650	797	0.60
Language Proficiency LEP	3,940	697.51	24.24	650	824	0.68
Language Proficiency FEP	4,167	719.44	32.36	650	850	0.86
Not Migrant	52,646	731.36	40.07	650	850	0.90
Migrant	158	702.00	29.78	650	772	0.78

*n-count less than 16

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	48,587	742.61	41.94	650	850	0.87
IEP	6,494	697.71	37.78	650	850	0.87
No Accommodation	51,253	740.32	43.02	650	850	0.88
Accommodation	3,828	697.09	35.39	650	850	0.85
Am. Indian/Alaska Native	364	714.09	38.58	650	812	0.87
Asian	1,955	753.23	43.22	650	850	0.88
Black	2,581	721.19	42.04	650	850	0.88
Hispanic	17,626	718.71	40.87	650	850	0.87
White	29,363	748.47	41.51	650	850	0.87
Hawaiian/Pacific Islander	186	721.17	42.96	650	839	0.89
Two or More Races	2,999	744.81	44.05	650	850	0.88
Missing	*	*	*	*	*	*
No Economic Disadvantage	36,667	747.53	42.22	650	850	0.87
Economic Disadvantage	18,414	716.97	40.02	650	850	0.87
Female	27,216	740.56	44.42	650	850	0.88
Male	27,865	734.15	43.21	650	850	0.88
Language Proficiency NA	46,663	741.97	43.08	650	850	0.88
Language Proficiency NEP	1,956	680.91	24.88	650	798	0.71
Language Proficiency LEP	5,426	713.22	34.65	650	850	0.83
Language Proficiency FEP	1,036	760.51	31.76	652	850	0.73
Not Migrant	54,916	737.42	43.91	650	850	0.88
Migrant	165	703.71	37.41	650	803	0.85

*n-count less than 16

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,108	745.58	33.61	650	850	0.87
IEP	6,638	702.86	34.20	650	831	0.89
No Accommodation	51,420	743.41	35.06	650	850	0.88
Accommodation	4,326	705.92	34.35	650	828	0.89
Am. Indian/Alaska Native	364	720.71	37.50	650	826	0.91
Asian	1,887	752.05	34.98	650	850	0.87
Black	2,513	724.67	35.36	650	828	0.89
Hispanic	18,286	725.30	35.03	650	850	0.88
White	29,602	750.31	33.61	650	850	0.86
Hawaiian/Pacific Islander	166	728.67	35.28	650	814	0.88
Two or More Races	2,922	745.49	35.59	650	850	0.88
Missing	*	*	*	*	*	*
No Economic Disadvantage	36,866	749.32	34.14	650	850	0.87
Economic Disadvantage	18,880	723.28	34.51	650	842	0.88
Female	27,174	743.91	36.39	650	850	0.88
Male	28,572	737.25	36.15	650	850	0.89

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Language Proficiency NA	46,657	744.74	35.33	650	850	0.88
Language Proficiency NEP	1,466	685.43	23.62	650	797	0.77
Language Proficiency LEP	5,570	715.62	28.39	650	822	0.83
Language Proficiency FEP	2,053	750.98	26.44	650	850	0.78
Not Migrant	55,592	740.58	36.40	650	850	0.89
Migrant	154	710.62	28.06	650	774	0.79

*n-count less than 16

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	50,416	749.61	31.06	650	850	0.95
IEP	6,946	714.02	27.49	650	842	0.93
No Accommodation	52,562	747.82	31.92	650	850	0.95
Accommodation	4,800	717.64	29.09	650	850	0.93
Am. Indian/Alaska Native	352	729.54	31.64	650	827	0.95
Asian	1,933	755.52	32.26	650	850	0.96
Black	2,516	732.01	30.37	650	842	0.95
Hispanic	19,958	731.58	29.76	650	850	0.94
White	29,480	754.77	31.14	650	850	0.96
Hawaiian/Pacific Islander	183	730.69	30.47	650	811	0.95
Two or More Races	2,937	750.98	33.13	650	850	0.96
Missing	*	*	*	*	*	*
No Economic Disadvantage	37,825	752.95	31.73	650	850	0.96
Economic Disadvantage	19,537	730.49	29.51	650	836	0.94
Female	28,094	749.11	33.03	650	850	0.95
Male	29,268	741.64	32.10	650	850	0.95
Language Proficiency NA	46,746	749.51	32.15	650	850	0.95
Language Proficiency NEP	1,130	699.31	20.49	650	772	0.89
Language Proficiency LEP	5,783	719.08	23.84	650	813	0.92
Language Proficiency FEP	3,703	747.11	25.07	650	845	0.94
Not Migrant	57,192	745.38	32.75	650	850	0.95
Migrant	170	717.35	28.41	650	781	0.91

*n-count less than 16

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,488	746.36	31.99	650	850	0.86
IEP	6,472	707.59	29.90	650	842	0.85
No Accommodation	51,549	744.44	33.05	650	850	0.87
Accommodation	4,411	711.91	31.72	650	829	0.87
Am. Indian/Alaska Native	363	723.67	33.76	650	834	0.88
Asian	1,730	757.61	33.88	650	850	0.87
Black	2,508	728.19	32.75	650	844	0.87

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Hispanic	19,956	727.13	31.64	650	845	0.86
White	28,507	752.19	31.43	650	850	0.85
Hawaiian/Pacific Islander	144	723.07	33.72	650	792	0.85
Two or More Races	2,746	748.11	33.56	650	850	0.87
Missing	*	*	*	*	*	*
No Economic Disadvantage	36,662	750.35	32.45	650	850	0.86
Economic Disadvantage	19,298	725.79	31.22	650	850	0.85
Female	27,353	746.17	33.87	650	850	0.87
Male	28,607	737.77	33.79	650	850	0.88
Language Proficiency NA	45,488	746.59	33.12	650	850	0.87
Language Proficiency NEP	928	690.72	23.81	650	798	0.73
Language Proficiency LEP	4,377	709.44	24.74	650	793	0.76
Language Proficiency FEP	5,167	737.09	26.99	650	850	0.81
Not Migrant	55,799	741.95	34.08	650	850	0.87
Migrant	161	718.92	29.97	650	795	0.84

*n-count less than 16

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,224	745.23	35.62	650	850	0.88
IEP	6,041	704.08	28.05	650	850	0.82
No Accommodation	50,867	743.48	36.32	650	850	0.88
Accommodation	4,398	708.90	31.62	650	850	0.87
Am. Indian/Alaska Native	365	724.58	33.35	650	830	0.87
Asian	1,751	759.15	37.22	650	850	0.88
Black	2,437	727.92	35.05	650	846	0.88
Hispanic	20,235	725.77	33.86	650	850	0.86
White	27,769	751.28	35.35	650	850	0.87
Hawaiian/Pacific Islander	163	727.75	37.29	650	829	0.89
Two or More Races	2,538	747.43	37.26	650	850	0.89
Missing	*	*	*	*	*	*
No Economic Disadvantage	36,531	749.23	36.12	650	850	0.88
Economic Disadvantage	18,734	724.17	33.39	650	841	0.86
Female	26,616	746.13	36.98	650	850	0.88
Male	28,649	735.72	36.63	650	850	0.89
Language Proficiency NA	44,856	745.73	36.39	650	850	0.88
Language Proficiency NEP	1,052	688.58	21.19	650	780	0.64
Language Proficiency LEP	4,024	707.25	25.80	650	829	0.74
Language Proficiency FEP	5,333	734.24	30.83	650	850	0.83
Not Migrant	55,104	740.81	37.16	650	850	0.88
Migrant	161	715.41	30.31	650	799	0.80

*n-count less than 16

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	47,217	747.07	39.36	650	850	0.95
IEP	5,510	702.28	30.08	650	838	0.93
No Accommodation	48,507	745.22	40.09	650	850	0.95
Accommodation	4,220	709.86	35.09	650	850	0.93
Am. Indian/Alaska Native	316	721.36	37.49	650	848	0.94
Asian	1,708	762.72	40.93	650	850	0.95
Black	2,291	729.90	38.45	650	850	0.95
Hispanic	19,603	727.04	37.62	650	850	0.93
White	26,270	753.29	39.20	650	850	0.95
Hawaiian/Pacific Islander	158	727.51	38.74	650	829	0.92
Two or More Races	2,375	749.66	40.48	650	850	0.95
Missing	*	*	*	*	*	*
No Economic Disadvantage	34,904	750.71	40.11	650	850	0.95
Economic Disadvantage	17,823	726.09	37.25	650	850	0.93
Female	25,104	751.28	40.13	650	850	0.94
Male	27,623	734.31	39.83	650	850	0.95
Language Proficiency NA	43,574	747.75	40.09	650	850	0.95
Language Proficiency NEP	1,032	684.77	23.25	650	785	0.89
Language Proficiency LEP	3,950	706.90	27.87	650	823	0.92
Language Proficiency FEP	4,171	734.22	32.60	650	850	0.93
Not Migrant	52,569	742.48	40.84	650	850	0.95
Migrant	158	712.70	34.45	650	790	0.91

*n-count less than 16

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	1,185	727.75	26.43	650	809	0.89
IEP	116	703.02	21.20	650	776	0.84
No Accommodation	1,151	727.54	26.49	650	809	0.89
Accommodation	150	710.23	25.46	650	781	0.88
Am. Indian/Alaska Native	*	*	*	*	*	*
Asian	*	*	*	*	*	*
Black	*	*	*	*	*	*
Hispanic	1,291	725.53	26.98	650	809	0.89
White	*	*	*	*	*	*
Hawaiian/Pacific Islander	*	*	*	*	*	*
Two or More Races	*	*	*	*	*	*
Missing	*	*	*	*	*	*
No Economic Disadvantage	394	721.45	26.84	650	793	0.89
Economic Disadvantage	907	727.32	26.81	650	809	0.89
Female	685	729.26	26.69	650	800	0.89
Male	616	721.41	26.63	650	809	0.89

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Language Proficiency NA	*	*	*	*	*	*
Language Proficiency NEP	558	716.82	26.30	650	800	0.89
Language Proficiency LEP	743	732.09	25.54	650	809	0.89
Language Proficiency FEP	*	*	*	*	*	*
Not Migrant	1,295	725.51	26.97	650	809	0.89
Migrant	*	*	*	*	*	*

*n-count less than 16

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

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	1,059	727.66	21.45	650	788	0.84
IEP	92	708.70	20.18	661	772	0.78
No Accommodation	986	727.56	21.70	650	788	0.84
Accommodation	165	717.72	21.62	650	770	0.84
Am. Indian/Alaska Native	*	*	*	*	*	*
Asian	*	*	*	*	*	*
Black	*	*	*	*	*	*
Hispanic	1,140	726.19	21.97	650	788	0.84
White	*	*	*	*	*	*
Hawaiian/Pacific Islander	*	*	*	*	*	*
Two or More Races	*	*	*	*	*	*
Missing	*	*	*	*	*	*
No Economic Disadvantage	404	724.06	23.08	650	788	0.85
Economic Disadvantage	747	727.28	21.25	650	788	0.84
Female	544	731.24	20.32	672	788	0.83
Male	607	721.59	22.38	650	788	0.85
Language Proficiency NA	*	*	*	*	*	*
Language Proficiency NEP	466	719.96	22.17	650	778	0.84
Language Proficiency LEP	685	730.36	20.80	661	788	0.84
Language Proficiency FEP	*	*	*	*	*	*
Not Migrant	1,150	726.13	21.96	650	788	0.84
Migrant	*	*	*	*	*	*

*n-count less than 16

Appendix I: Summary Statistics for Points Earned by Subclaim

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

Subclaim	Grade	Mean	SD	Min.	Max.	Max. Points Possible	Average % Correct
Subclaim A	3	11.3	5.3	0	22	22	51.27
	4	10.7	6.1	0	24	24	44.49
	5	11.2	5.9	0	23	23	48.54
	6	6.5	4.7	0	20	20	32.56
	7	6.7	4.6	0	23	23	29.05
	8	8.9	5.1	0	24	24	37.16
Subclaim B	3	4.9	2.5	0	9	9	54.84
	4	2.6	1.8	0	7	7	36.76
	5	3.4	2.3	0	8	8	43.16
	6	4.6	2.4	0	11	11	41.34
	7	2.8	1.8	0	8	8	35.44
	8	2.8	1.9	0	7	7	39.45
Subclaim C	3	3.5	2.7	0	11	11	31.38
	4	3.3	3.2	0	11	11	30.06
	5	2.4	2.6	0	11	11	21.86
	6	3.2	3.0	0	11	11	29.42
	7	2.5	2.8	0	11	11	22.96
	8	1.8	2.3	0	11	11	16.61
Subclaim D	3	3.0	2.3	0	9	9	3.08
	4	2.2	2.6	0	9	9	2.16
	5	2.2	2.5	0	9	9	2.31
	6	1.9	2.3	0	9	9	2.04
	7	2.1	2.2	0	9	9	2.20
	8	1.8	2.3	0	9	9	1.74

Note. One item was removed from scoring in Grade 6.

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

Subclaim	Grade	Mean	SD	Min.	Max.	Max. Points Possible	Average % Correct
RL	3	6.3	3.6	0	17	17	36.83
	4	8.6	3.9	0	18	18	47.88
	5	7.1	4.1	0	18	18	39.49
	6	8.1	4.5	0	18	18	45.23
	7	6.7	4.0	0	18	18	37.09
	8	8.3	4.2	0	18	18	46.20
RI	3	7.8	3.6	0	14	14	55.96
	4	7.6	4.3	0	18	18	42.44
	5	6.1	3.3	0	18	18	34.15
	6	9.3	4.3	0	22	22	42.22
	7	8.1	4.5	0	22	22	36.69
	8	9.3	4.7	0	22	22	42.27
RV	3	5.9	2.7	0	10	10	59.28
	4	6.5	2.6	0	10	10	64.63
	5	4.5	2.0	0	8	8	56.85
	6	4.6	2.2	0	8	8	57.07
	7	5.1	2.7	0	10	10	50.63
	8	4.4	2.7	0	10	10	44.12
WE (unweighted)	3	1.2	1.1	0	6	6	20.58
	4	1.7	1.4	0	7	7	24.38
	5	1.3	1.3	0	7	7	18.35
	6	1.6	1.6	0	8	8	20.37
	7	1.9	1.8	0	8	8	23.98
	8	2.6	2.1	0	8	8	32.68
WKL	3	1.2	1.2	0	6	6	19.60
	4	1.5	1.4	0	6	6	24.86
	5	1.2	1.3	0	6	6	20.16
	6	1.6	1.6	0	6	6	26.31
	7	1.8	1.7	0	6	6	30.05
	8	2.4	2.0	0	6	6	39.99

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 I.3. Points Earned Summary by Subclaim—CSLA

Subclaim	Grade	Mean	SD	Min.	Max.	Max. Points Possible	Average % Correct
RL	3	6.3	4.2	0	17	17	37.14
	4	7.7	4.0	0	18	20	38.46
RI	3	4.4	3.1	0	14	14	31.06
	4	4.4	2.9	0	15	18	24.66
RV	3	3.8	2.8	0	10	10	37.59
	4	3.3	2.4	0	8	8	41.37
WE (unweighted)	3	1.4	1.4	0	6	6	23.57
	4	2.0	1.7	0	7	7	28.88
WKL	3	1.9	1.8	0	6	6	30.83
	4	2.3	1.8	0	6	6	38.50

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.

Appendix J: Classical Item-Level Statistics

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

Item	Omit %	*P*-value	Item–Total Correlation
1	0.90	0.73	0.50
2	0.40	0.68	0.59
3	0.40	0.45	0.61
4	3.40	0.36	0.65
5	1.20	0.41	0.57
6	2.30	0.46	0.40
7	0.50	0.55	0.56
8	0.40	0.27	0.43
9	0.60	0.78	0.35
10	3.50	0.39	0.61
11	1.00	0.22	0.39
12	0.30	0.62	0.51
13	0.40	0.51	0.67
14	0.40	0.44	0.66
15	1.10	0.56	0.64
16	0.30	0.80	0.45
17	0.10	0.77	0.50
18	3.00	0.54	0.43
19	0.80	0.76	0.52
20	1.40	0.51	0.47
21	0.00	0.82	0.46
22	0.20	0.48	0.44
23	0.70	0.60	0.60
24	0.30	0.80	0.41
25	0.30	0.56	0.59
26	0.10	0.35	0.58
27	2.60	0.34	0.70
28	2.00	0.46	0.45
29	0.80	0.38	0.75
30	0.40	0.33	0.72
31	0.00	0.32	0.78

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

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
1	3	3.00	36.11	18.21	31.14	11.53	–	0.38	0.61
2	4	1.40	33.88	39.30	16.68	7.27	1.51	0.25	0.65

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

Item	Omit %	*P*-value	Item–Total Correlation
1	1.40	0.20	0.51
2	0.10	0.52	0.53
3	1.40	0.39	0.60
4	0.50	0.17	0.56
5	1.20	0.16	0.57
6	0.80	0.62	0.39
7	0.90	0.37	0.48
8	3.40	0.28	0.64
9	1.60	0.41	0.42
10	0.20	0.63	0.53
11	4.00	0.29	0.57
12	0.10	0.41	0.45
13	0.10	0.63	0.51
14	0.60	0.61	0.45
15	0.00	0.71	0.56
16	0.50	0.44	0.52
17	0.10	0.43	0.37
18	3.70	0.44	0.55
19	1.70	0.26	0.40
20	1.90	0.28	0.69
21	1.30	0.53	0.50
22	0.30	0.52	0.66
23	1.40	0.36	0.67
24	0.70	0.34	0.52
25	0.10	0.66	0.61
26	0.40	0.40	0.74

Table J.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	2.40	61.37	12.16	10.25	13.82	–	–	–	0.25	0.67
2	3	3.40	74.21	7.39	5.43	9.53	–	–	–	0.16	0.63
3	4	6.60	45.40	18.70	13.60	8.30	7.40	–	–	0.25	0.75
4	6	1.10	43.28	10.32	14.67	6.71	8.96	6.92	8.076	0.30	0.78

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

Item	Omit %	*P*-value	Item–Total Correlation
1	0.10	0.53	0.54
2	1.40	0.33	0.45
3	0.90	0.63	0.52
4	2.40	0.36	0.60
5	0.30	0.34	0.64
6	0.10	0.83	0.29
7	1.00	0.48	0.63
8	0.70	0.37	0.57
9	0.70	0.43	0.57
10	0.30	0.33	0.49
11	0.30	0.65	0.58
12	0.50	0.72	0.32
13	2.30	0.62	0.54
14	0.20	0.62	0.42
15	0.20	0.27	0.40
16	0.10	0.72	0.53
17	0.70	0.52	0.48
18	3.80	0.44	0.37
19	0.20	0.44	0.52
20	0.10	0.56	0.71
21	3.00	0.38	0.60
22	0.50	0.53	0.62
23	1.40	0.36	0.76
24	0.10	0.44	0.72
25	0.80	0.28	0.61
26	1.90	0.27	0.77
27	0.20	0.23	0.77

Table J.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	3	6.00	54.74	25.55	10.04	3.67	–	–	–	0.19	0.68
2	4	2.50	52.51	15.69	14.61	6.12	8.58	–	–	0.24	0.72
3	6	2.10	49.00	5.70	18.10	8.50	6.40	7.30	2.80	0.24	0.80

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

Item	Omit %	*P*-value	Item–Total Correlation
1	0.20	0.57	0.52
2	0.30	0.21	0.46
3	0.40	0.58	0.42
4	0.70	0.09	0.38
5	2.00	0.31	0.62
6	0.30	0.24	0.52
7	0.30	0.26	0.37
8	0.90	0.23	0.56
9	3.50	0.26	0.55
10	3.40	0.29	0.65
11	2.30	0.23	0.61
12	1.00	0.50	0.49
13	0.60	0.39	0.16
14	0.80	0.29	0.52
15	0.20	0.48	0.39
16	0.40	0.27	0.21
17	1.00	0.26	0.28
18	0.50	0.60	0.48
19	0.20	0.32	0.53
20	0.90	0.38	0.71
21	1.70	0.20	0.65
22	0.10	0.58	0.55
23	0.00	0.44	0.38
24	0.40	0.37	0.66

Table J.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	1.00	37.07	13.53	29.25	19.20	–	–	–	0.43	0.72
2	3	1.00	48.26	19.26	23.55	7.91	–	–	–	0.30	0.69
3	4	2.70	68.71	12.66	6.42	6.68	2.80	–	–	0.14	0.70
4	4	2.10	42.08	12.46	16.53	18.43	8.41	–	–	0.34	0.74
5	6	3.30	60.20	9.40	7.30	7.60	6.10	4.10	2.10	0.17	0.79

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

Item	Omit %	*P*-value	Item–Total Correlation
1	0.70	0.43	0.57
2	0.20	0.20	0.49
3	0.10	0.48	0.55
4	0.30	0.18	0.54
5	1.40	0.13	0.57
6	0.80	0.31	0.39
7	0.40	0.24	0.67
8	0.40	0.49	0.34
9	0.10	0.65	0.50
10	0.40	0.44	0.44
11	0.10	0.19	0.36
12	2.80	0.58	0.36
13	0.60	0.30	0.25
14	0.50	0.20	0.42
15	0.40	0.22	0.47
16	0.20	0.41	0.24
17	0.20	0.54	0.42
18	0.10	0.20	0.25
19	0.70	0.55	0.45
20	1.10	0.13	0.42
21	1.10	0.15	0.67
22	0.30	0.38	0.45
23	2.30	0.12	0.60
24	0.70	0.23	0.70
25	0.30	0.29	0.70

Table J.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	4.50	56.07	13.70	15.56	10.13	–	–	–	0.25	0.77
2	3	3.30	55.41	30.47	9.01	1.85	–	–	–	0.18	0.65
3	4	3.00	56.73	18.46	6.31	7.30	8.25	–	–	0.22	0.77
4	6	2.50	32.10	30.20	11.00	7.80	8.20	5.30	2.90	0.25	0.82

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

Item	Omit %	*P*-value	Item–Total Correlation
1	1.40	0.15	0.63
2	0.60	0.38	0.60
3	1.30	0.26	0.60
4	2.30	0.13	0.60
5	1.40	0.33	0.50
6	0.10	0.32	0.49
7	0.70	0.26	0.57
8	0.10	0.43	0.34
9	0.10	0.40	0.30
10	0.10	0.34	0.44
11	0.10	0.28	0.40
12	0.10	0.85	0.32
13	0.10	0.70	0.40
14	0.70	0.26	0.44
15	0.40	0.66	0.45
16	0.10	0.53	0.36
17	0.20	0.48	0.28
18	0.20	0.50	0.49
19	1.70	0.25	0.69
20	0.10	0.24	0.46
21	1.20	0.37	0.66
22	0.10	0.45	0.52
23	0.60	0.32	0.72
24	0.70	0.24	0.76

Table J.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	3	3.60	68.57	7.74	12.19	7.94	–	–	–	0.19	0.76
2	3	1.70	61.56	16.26	8.09	12.37	–	–	–	0.23	0.70
3	4	5.70	78.05	8.16	3.92	3.17	1.04	–	–	0.07	0.66
4	6	3.80	50.40	18.80	13.60	3.50	3.10	4.10	2.80	0.18	0.82

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

Item	Omit %	*P*-value	Item–Total Correlation
1	0.80	0.23	0.40
2	0.80	0.36	0.40
3	0.60	0.30	0.33
4	1.60	0.54	0.55
5	1.20	0.45	0.56
6	0.10	0.81	0.53
7	0.10	0.65	0.61
8	0.10	0.66	0.56
9	0.20	0.42	0.53
10	1.70	0.77	0.61
11	2.40	0.79	0.51
12	2.90	0.31	0.53
13	2.20	0.56	0.59
14	2.00	0.58	0.53
15	0.00	0.58	0.54
16	0.20	0.45	0.55
17	0.10	0.48	0.56
18	0.20	0.51	0.59
19	2.80	0.30	0.51

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

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
PCR 1 WE	3	1.40	40.17	49.00	8.93	0.54	–	0.23	0.69
PCR 1 WKL	3	1.40	49.69	36.65	11.63	0.67	–	0.21	0.63
PCR 2 WE	3	1.80	50.48	40.10	7.11	0.50	–	0.19	0.74
PCR 2 WKL	3	1.80	50.60	41.66	5.33	0.59	–	0.18	0.63

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

Item	Omit %	*P*-value	Item–Total Correlation
1	0.10	0.63	0.44
2	0.70	0.35	0.33
3	0.10	0.49	0.52
4	0.00	0.56	0.50
5	0.30	0.46	0.48
6	0.10	0.79	0.54
7	0.20	0.30	0.33
8	0.30	0.62	0.67
9	0.40	0.52	0.61
10	0.10	0.25	0.29
11	0.20	0.46	0.61
12	0.30	0.46	0.46
13	0.20	0.75	0.54
14	0.30	0.44	0.60
15	0.40	0.30	0.38
16	0.30	0.47	0.59
17	0.10	0.60	0.58
18	0.10	0.63	0.45
19	0.10	0.74	0.46
20	0.10	0.50	0.68
21	0.10	0.68	0.65

Table J.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	0.80	33.46	41.08	19.35	5.33	–	0.32	0.76
PCR 1 WKL	3	0.80	40.95	37.50	16.74	4.02	–	0.28	0.73
PCR 2 WE	4	1.10	38.52	49.50	9.32	1.33	0.27	0.18	0.73
PCR 2 WKL	3	1.10	47.80	39.07	10.09	1.98	–	0.22	0.68

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

Item	Omit %	*P*-value	Item–Total Correlation
1	0.20	0.77	0.50
2	0.10	0.60	0.57
3	0.30	0.48	0.52
4	0.40	0.47	0.49
5	0.30	0.33	0.36
6	0.30	0.25	0.42
7	0.30	0.65	0.64
8	0.10	0.40	0.45
9	0.10	0.29	0.52
10	0.10	0.39	0.39
11	0.10	0.30	0.39
12	0.20	0.31	0.46
13	0.10	0.36	0.32
14	0.20	0.54	0.31
15	0.10	0.50	0.41
16	0.10	0.21	0.33
17	0.40	0.49	0.55
18	0.10	0.46	0.45
19	0.30	0.36	0.46
20	0.60	0.46	0.49

Table J.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	0.90	50.83	30.18	14.19	3.94	–	0.24	0.77
PCR 1 WKL	3	0.90	55.92	27.73	12.05	3.45	–	0.21	0.72
PCR 2 WE	4	1.60	50.69	39.43	6.86	1.16	0.22	0.14	0.73
PCR 2 WKL	3	1.60	50.78	39.40	7.04	1.14	–	0.19	0.69

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

Item	Omit %	*P*-value	Item–Total Correlation
1	0.40	0.31	0.39
2	0.40	0.47	0.47
3	0.50	0.35	0.48
4	0.50	0.74	0.51
5	0.50	0.34	0.34
6	0.40	0.68	0.58
7	0.10	0.35	0.43
8	0.20	0.59	0.40
9	0.10	0.36	0.39
10	0.30	0.53	0.44
11	0.60	0.55	0.56
12	0.50	0.43	0.43
13	0.40	0.29	0.54
14	0.10	0.67	0.51
15	0.40	0.41	0.56
16	0.40	0.49	0.62
17	0.40	0.46	0.54
18	0.40	0.48	0.62
19	1.20	0.66	0.45
20	2.00	0.52	0.30
21	1.50	0.48	0.42
22	1.70	0.42	0.34

Table J.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	2.20	45.00	35.72	13.68	2.94	0.47	0.19	0.79
PCR 1 WKL	3	2.20	43.93	34.89	15.41	3.58	–	0.26	0.76
PCR 2 WE	4	2.00	41.08	35.15	16.77	4.00	0.95	0.21	0.78
PCR 2 WKL	3	2.00	46.44	31.02	15.78	4.72	–	0.26	0.76

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

Item	Omit %	*P*-value	Item–Total Correlation
1	0.50	0.58	0.49
2	0.60	0.31	0.40
3	0.70	0.32	0.51
4	0.60	0.42	0.50
5	0.60	0.43	0.54
6	0.70	0.43	0.38
7	0.60	0.51	0.63
8	0.50	0.67	0.57
9	0.50	0.34	0.40
10	0.50	0.45	0.55
11	0.50	0.44	0.39
12	0.40	0.56	0.45
13	1.00	0.37	0.46
14	1.80	0.32	0.50
15	1.80	0.24	0.52
16	1.30	0.42	0.54
17	0.00	0.25	0.35
18	0.20	0.24	0.42
19	0.10	0.54	0.41
20	0.20	0.44	0.43
21	0.20	0.56	0.58
22	0.20	0.35	0.40
23	0.30	0.32	0.51

Table J.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	3.00	43.26	37.50	12.64	3.14	0.44	0.19	0.82
PCR 1 WKL	3	3.00	40.34	41.31	12.21	3.12	–	0.25	0.79
PCR 2 WE	4	2.10	34.25	28.47	22.93	10.00	2.21	0.28	0.82
PCR 2 WKL	3	2.10	38.56	27.27	21.52	10.51	–	0.34	0.80

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

Item	Omit %	*P*-value	Item–Total Correlation
1	0.40	0.55	0.67
2	0.30	0.51	0.55
3	0.40	0.46	0.49
4	0.40	0.43	0.47
5	0.30	0.36	0.48
6	0.00	0.29	0.39
7	4.10	0.43	0.51
8	0.00	0.64	0.46
9	0.10	0.48	0.58
10	0.10	0.31	0.41
11	0.00	0.55	0.50
12	0.10	0.27	0.30
13	0.40	0.44	0.49
14	0.30	0.65	0.43
15	0.20	0.45	0.39
16	0.20	0.33	0.36
17	0.20	0.34	0.49
18	0.50	0.57	0.53
19	0.50	0.25	0.31
20	0.40	0.52	0.55
21	0.50	0.56	0.43
22	0.40	0.66	0.59
23	0.40	0.29	0.41

Table J.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	2.80	27.84	25.41	29.66	11.67	2.58	0.33	0.86
PCR 1 WKL	3	2.80	28.13	25.52	30.62	12.89	–	0.42	0.85
PCR 2 WE	4	2.20	30.22	26.38	26.16	11.75	3.27	0.32	0.86
PCR 2 WKL	3	2.20	37.60	22.98	23.97	13.22	–	0.37	0.84

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

Item	Omit %	*P*-value	Item–Total Correlation
1	4.50	0.42	0.60
2	5.10	0.32	0.56
3	4.90	0.41	0.61
4	5.20	0.39	0.56
5	1.30	0.33	0.44
6	1.10	0.30	0.47
7	1.40	0.43	0.62
8	0.50	0.34	0.57
9	0.80	0.43	0.39
10	0.50	0.41	0.54
11	1.50	0.38	0.43
12	0.80	0.51	0.50
13	1.10	0.48	0.55
14	1.50	0.38	0.56
15	12.90	0.38	0.59
16	15.80	0.29	0.51
17	15.30	0.31	0.45
18	15.00	0.26	0.50
19	16.40	0.16	0.26

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

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
PCR 1 WE	3	5.70	43.74	25.67	16.68	8.22	–	0.28	0.71
PCR 1 WKL	3	5.70	36.59	32.74	13.99	10.99	–	0.31	0.59
PCR 2 WE	3	5.10	48.19	37.28	7.84	1.54	–	0.19	0.73
PCR 2 WKL	3	5.10	35.67	37.20	11.84	10.15	–	0.30	0.59

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

Item	Omit %	*P*-value	Item–Total Correlation
1	0.80	0.38	0.49
2	0.40	0.40	0.60
3	1.00	0.43	0.45
4	0.80	0.40	0.51
5	1.10	0.27	0.48
6	1.00	0.17	0.17
7	1.00	0.62	0.57
8	0.90	0.45	0.33
9	0.90	0.32	0.45
10	1.00	0.63	0.41
11	3.60	0.15	0.12
12	3.70	0.27	0.29
13	3.60	0.17	0.24
14	3.00	0.41	0.45
15	3.00	0.23	0.33
16	3.60	0.45	0.49
17	3.70	0.26	0.41
18	1.00	0.47	0.50
19	1.30	0.26	0.33
20	1.60	0.18	0.25
21	1.30	0.26	0.27

Table J.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	3.20	33.19	23.89	31.19	7.73	0.78	0.28	0.78
PCR 1 WKL	3	3.20	34.93	35.71	16.33	9.82	–	0.33	0.60
PCR 2 WE	3	1.90	49.09	17.20	23.28	8.51	–	0.30	0.66
PCR 2 WKL	3	1.90	23.11	36.66	18.42	19.90	–	0.44	0.58

Appendix K: Scree Plots

Figure K.1. Scree Plot—Mathematics Grade 3

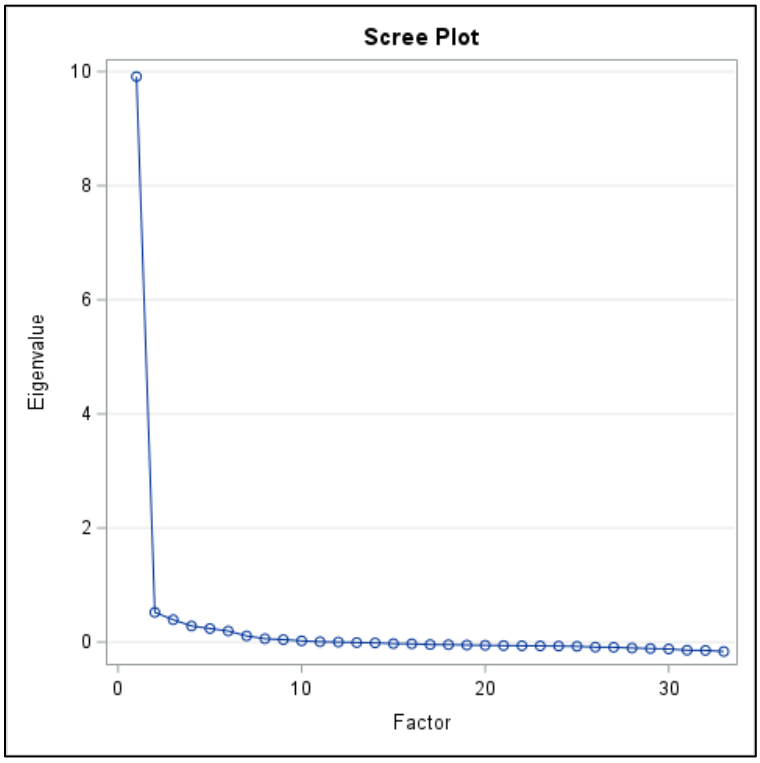


Figure K.2. Scree Plot—Mathematics Grade 4

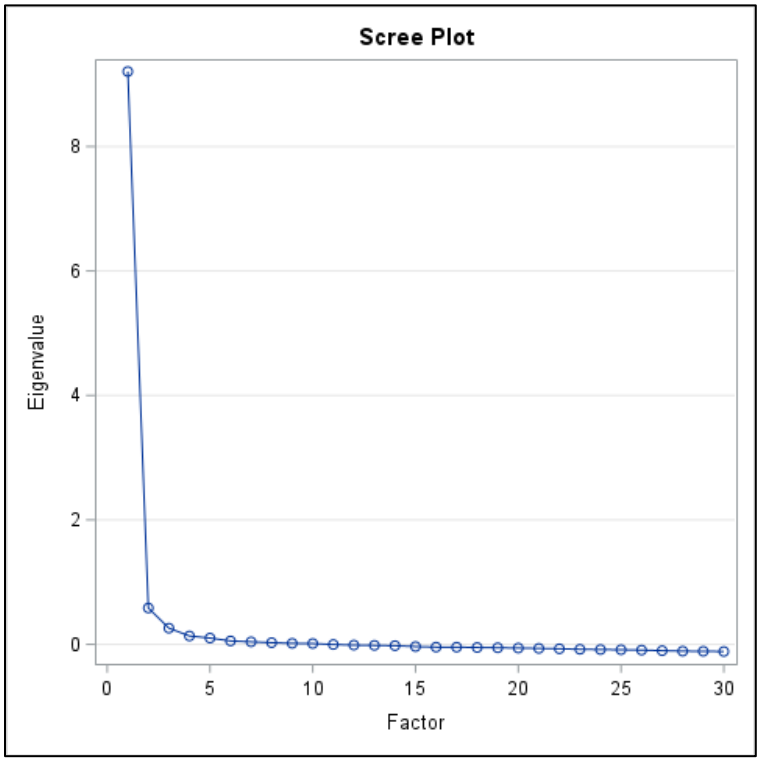


Figure K.3. Scree Plot—Mathematics Grade 5

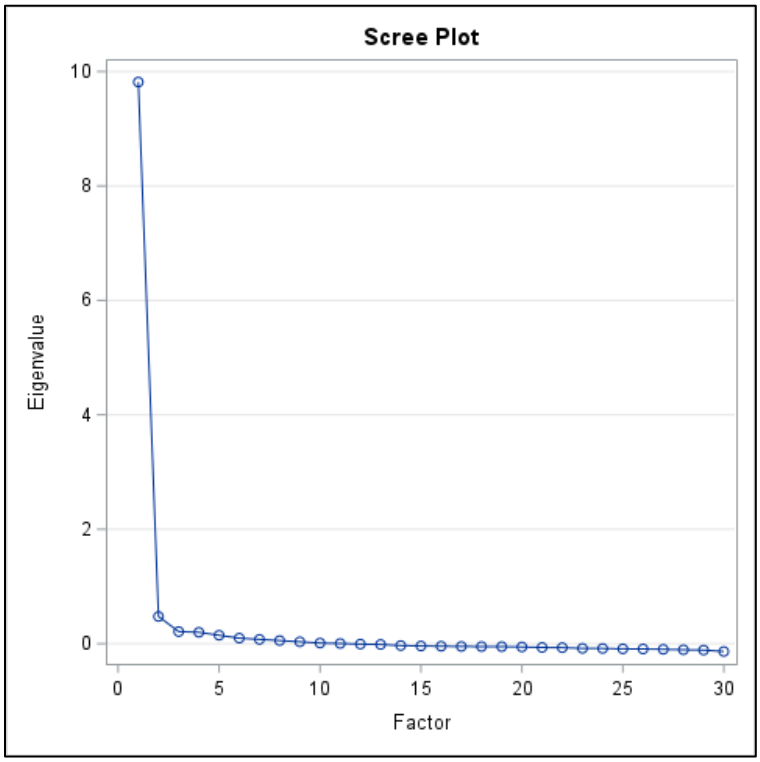


Figure K.4. Scree Plot—Mathematics Grade 6

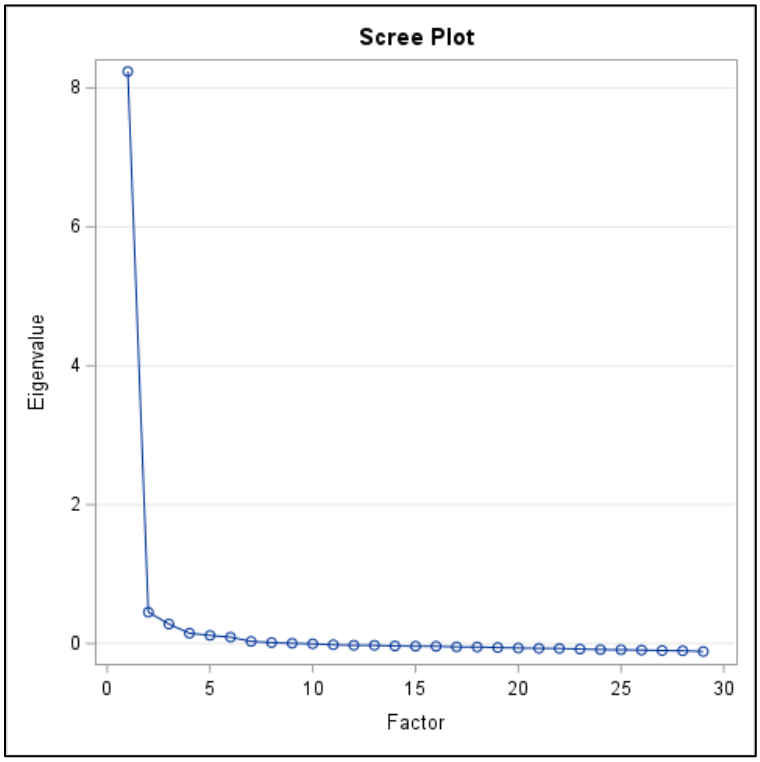


Figure K.5. Scree Plot—Mathematics Grade 7

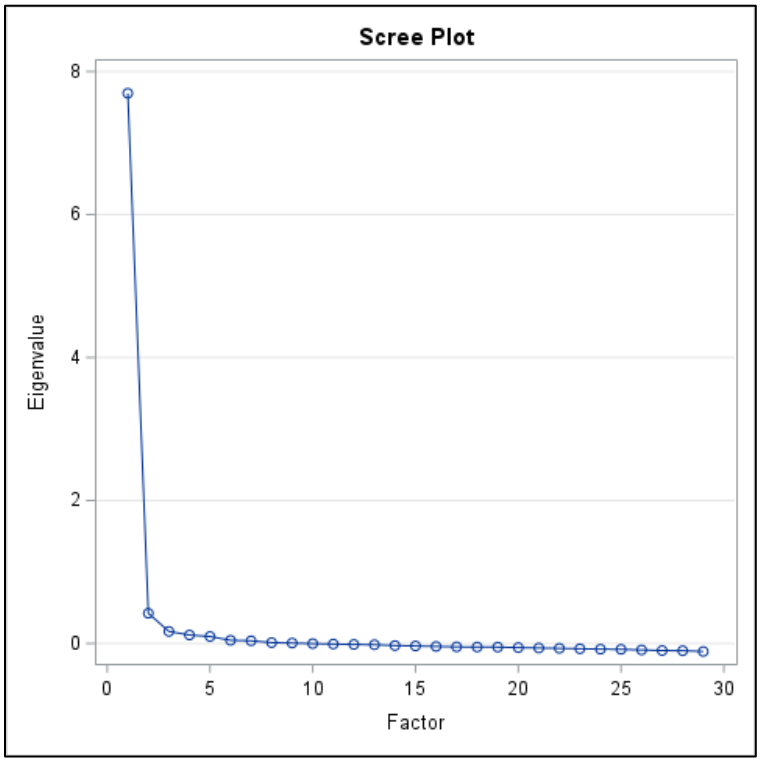


Figure K.6. Scree Plot—Mathematics Grade 8

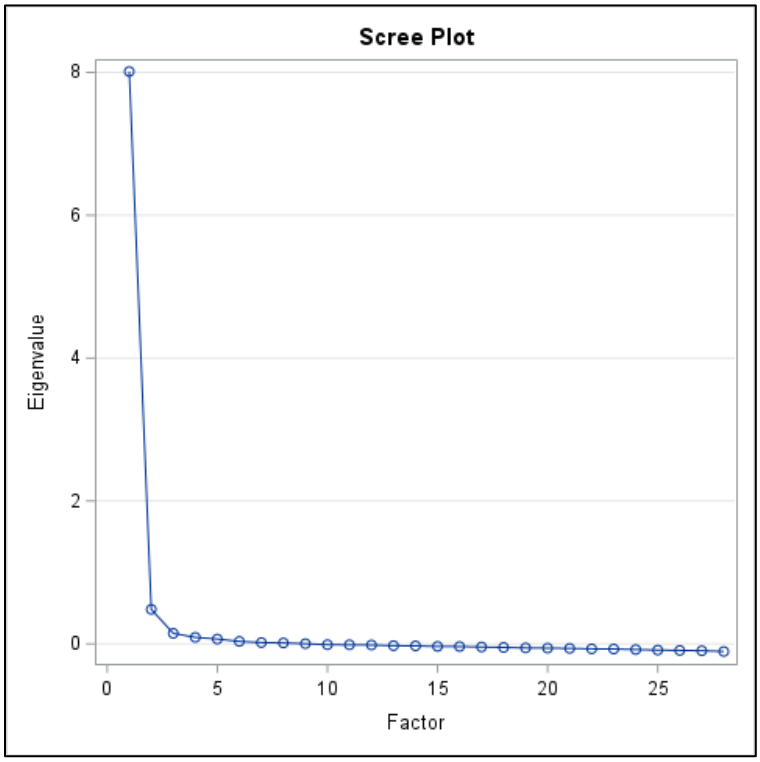


Figure K.7. Scree Plot—ELA Grade 3

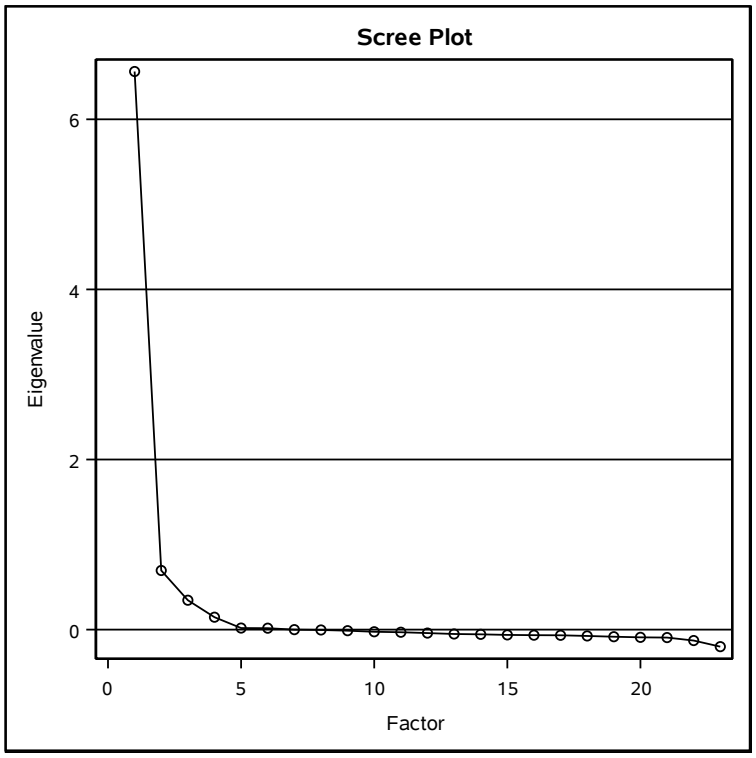


Figure K.8. Scree Plot—ELA Grade 4

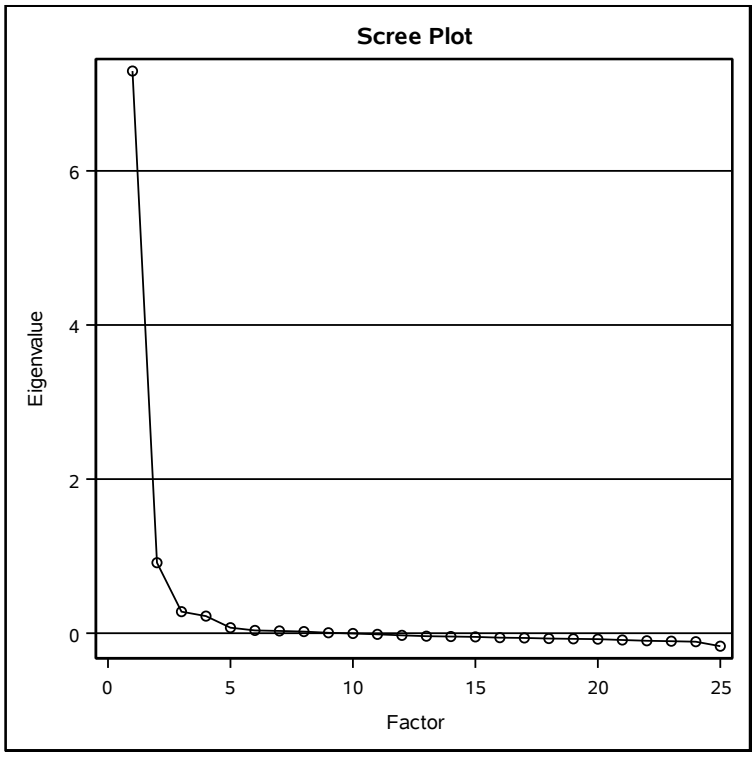


Figure K.9. Scree Plot—ELA Grade 5

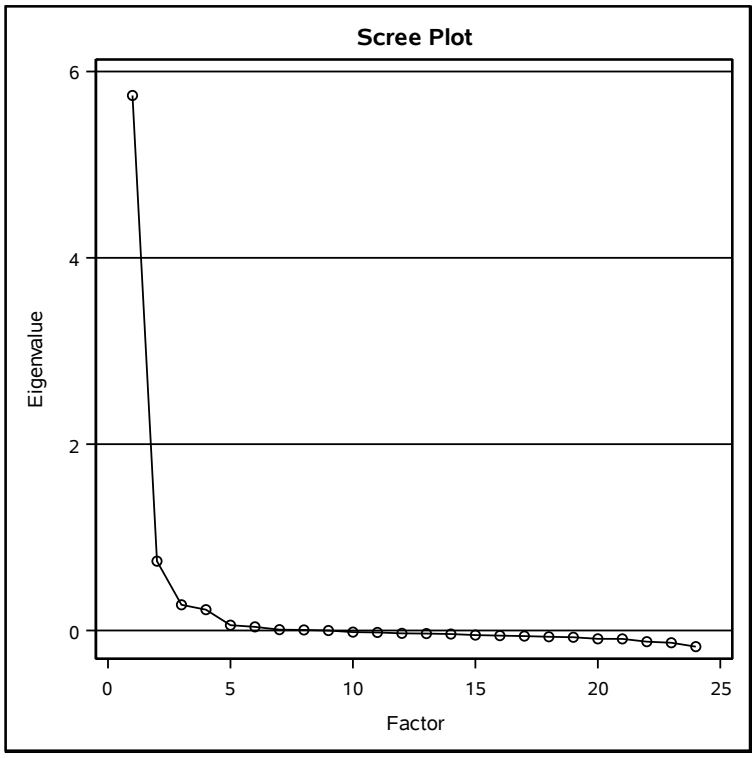


Figure K.10. Scree Plot—ELA Grade 6

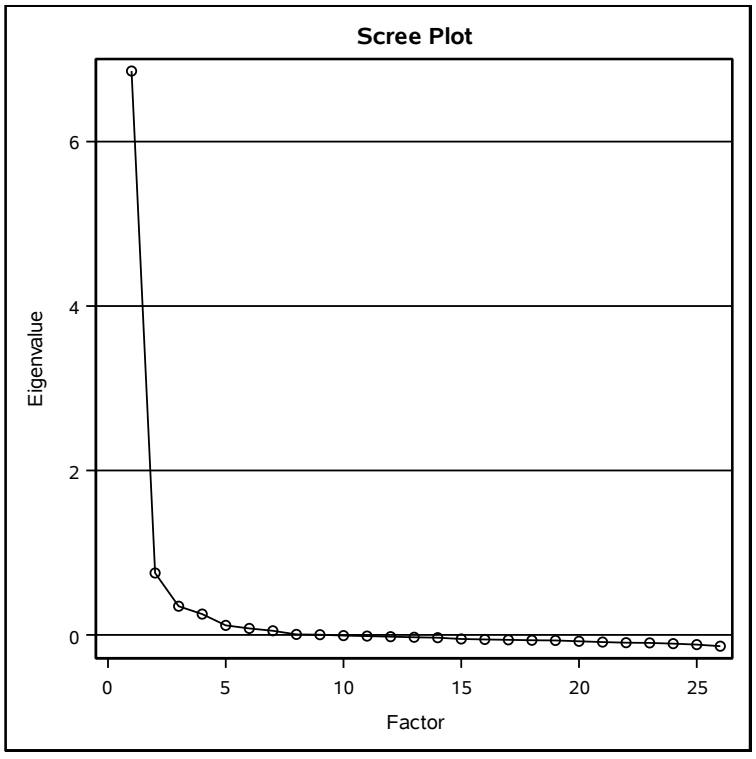


Figure K.11. Scree Plot—ELA Grade 7

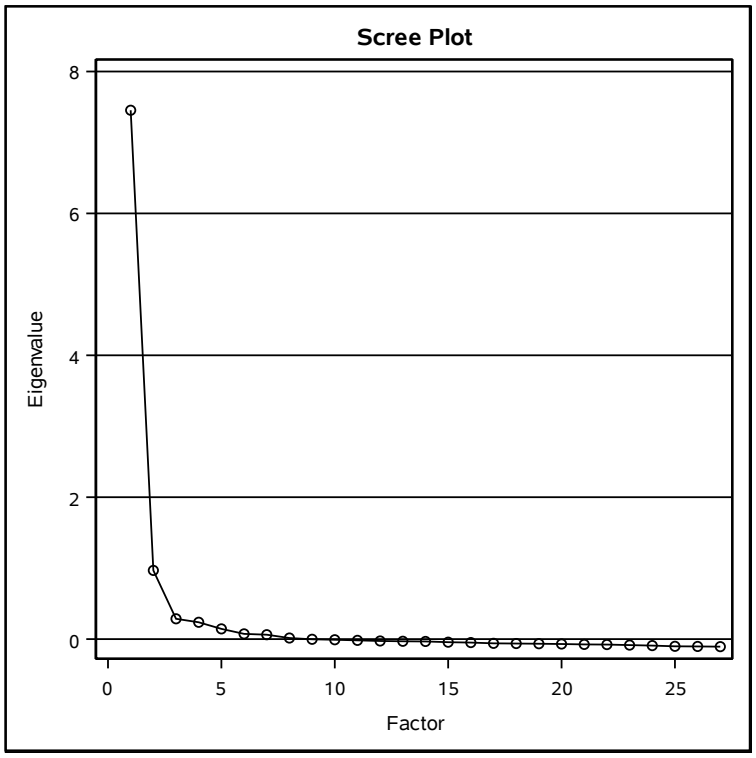


Figure K.12. Scree Plot—ELA Grade 8

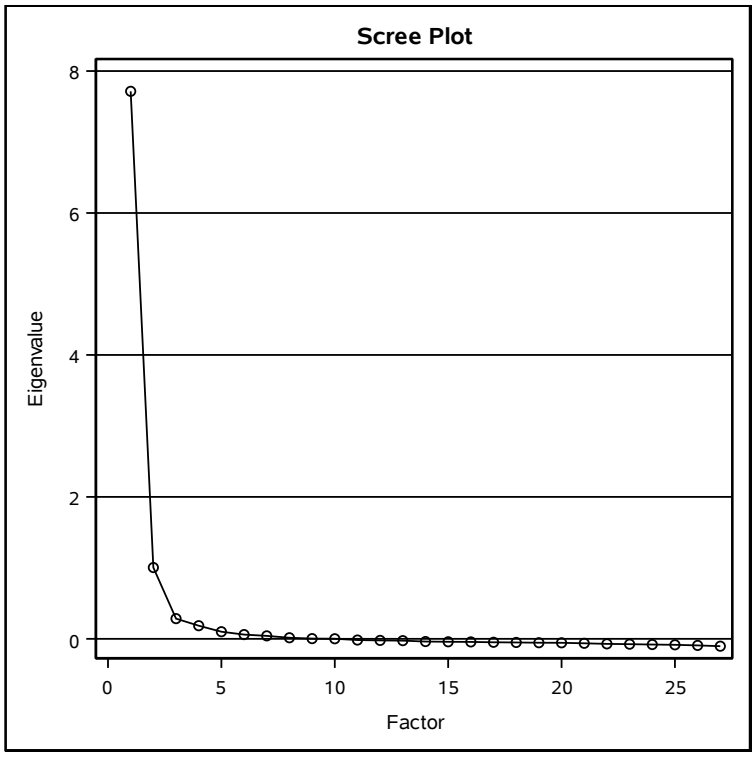


Figure K.13. Scree Plot—CSLA Grade 3

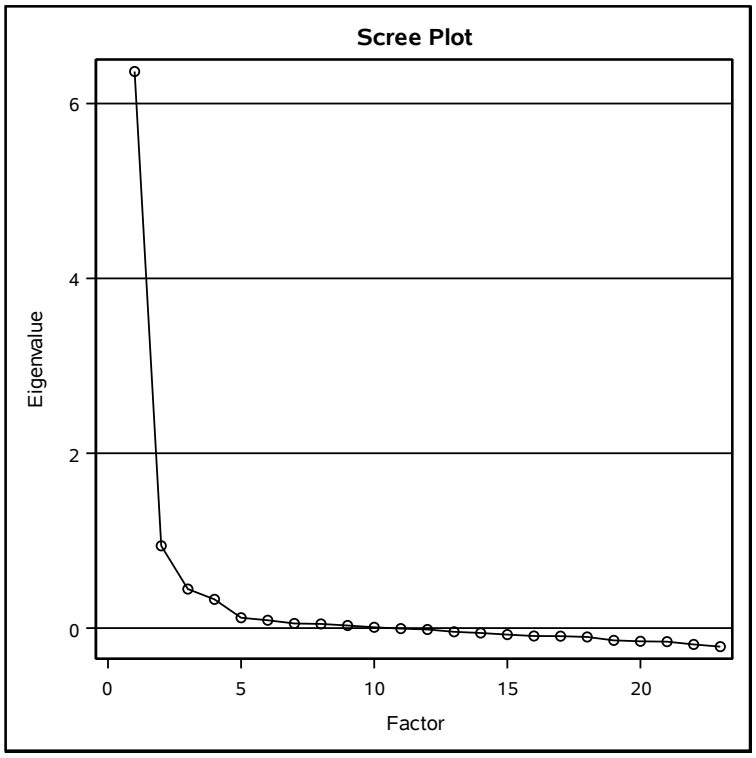
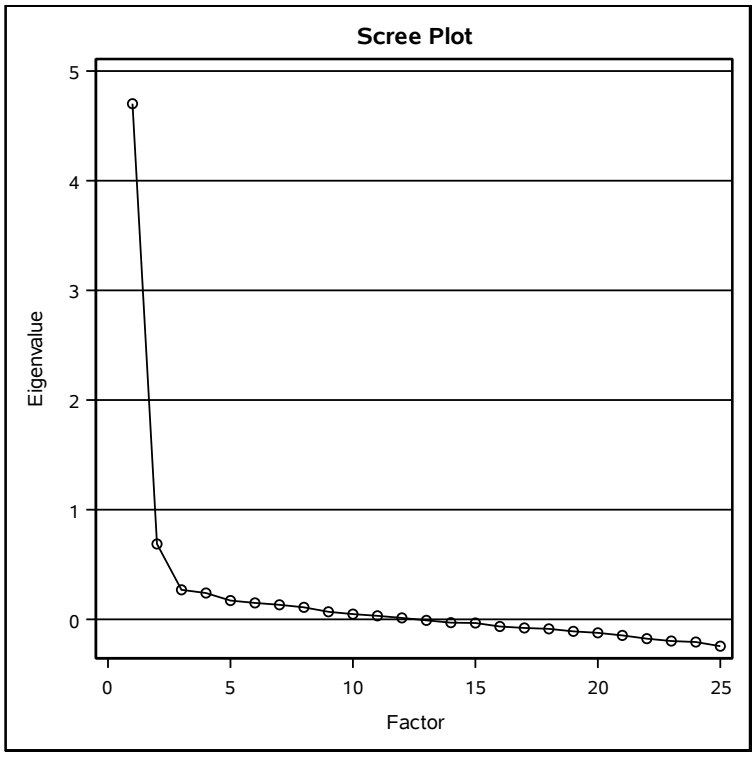


Figure K.14. Scree Plot—CSLA Grade 4



Appendix L: Mathematics Post-Equating Check TCCs

Figure L.1. Post-Equating Check TCCs—Mathematics Grade 3

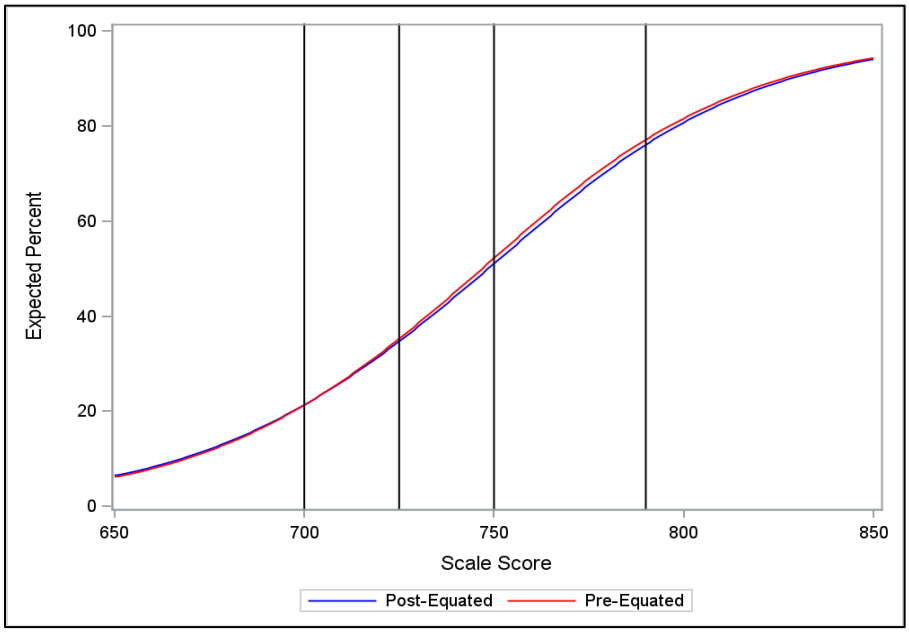


Figure L.2. Post-Equating Check TCCs—Mathematics Grade 4

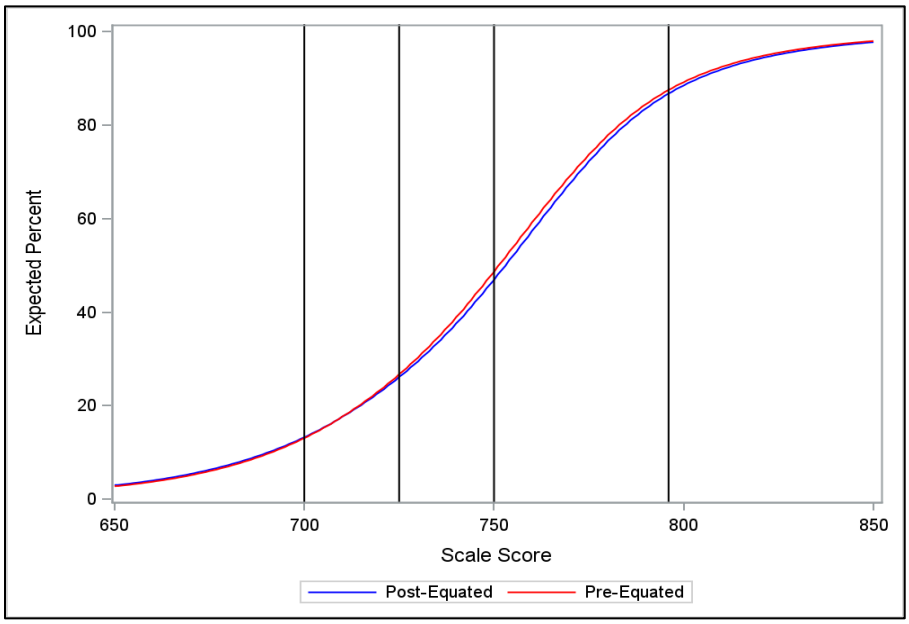


Figure L.3. Post-Equating Check TCCs—Mathematics Grade 5

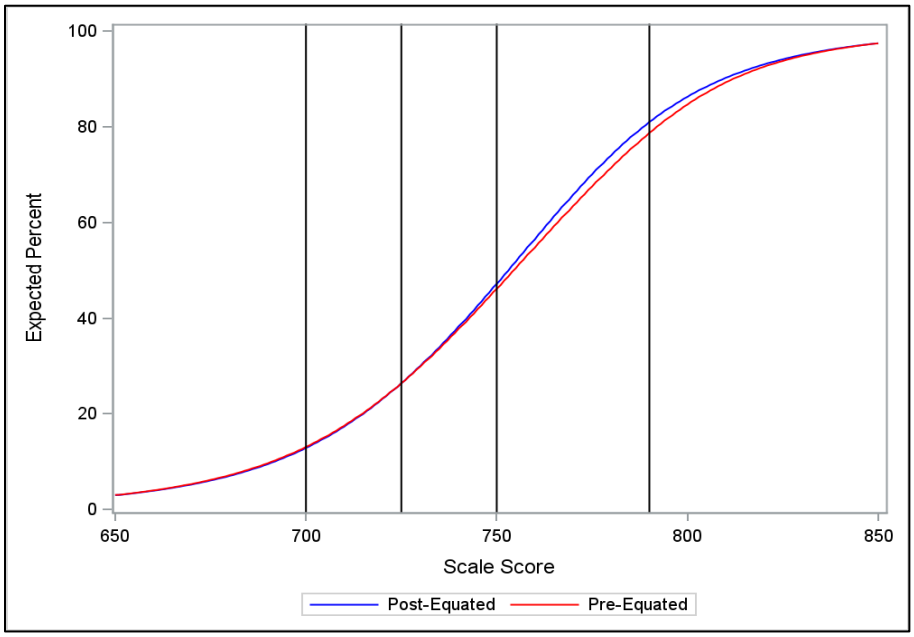


Figure L.4. Post-Equating Check TCCs—Mathematics Grade 6

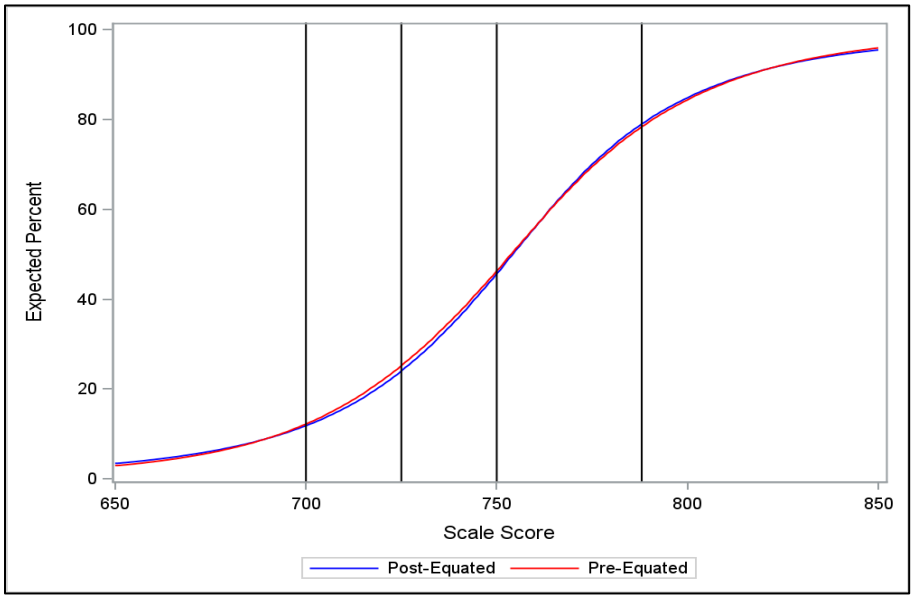


Figure L.5. Post-Equating Check TCCs—Mathematics Grade 7

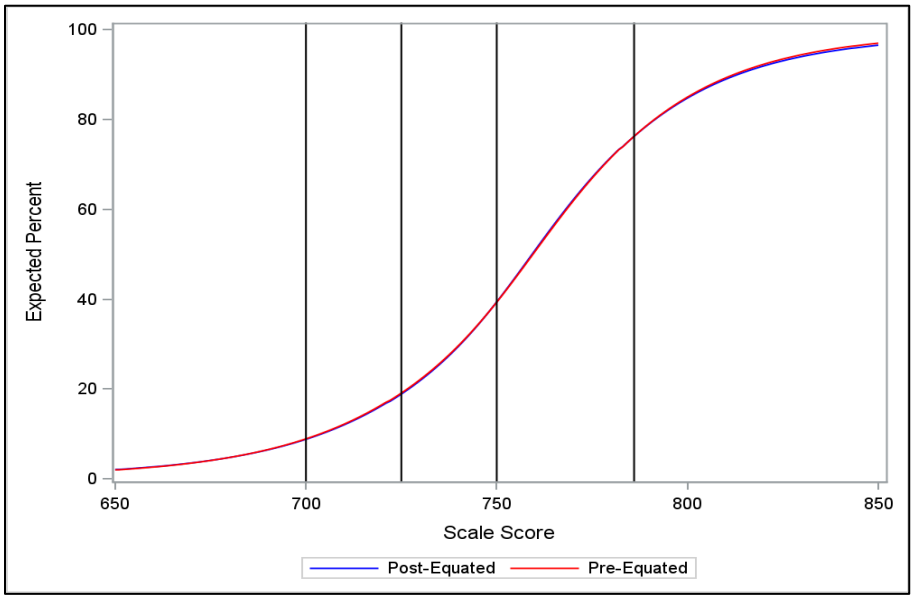
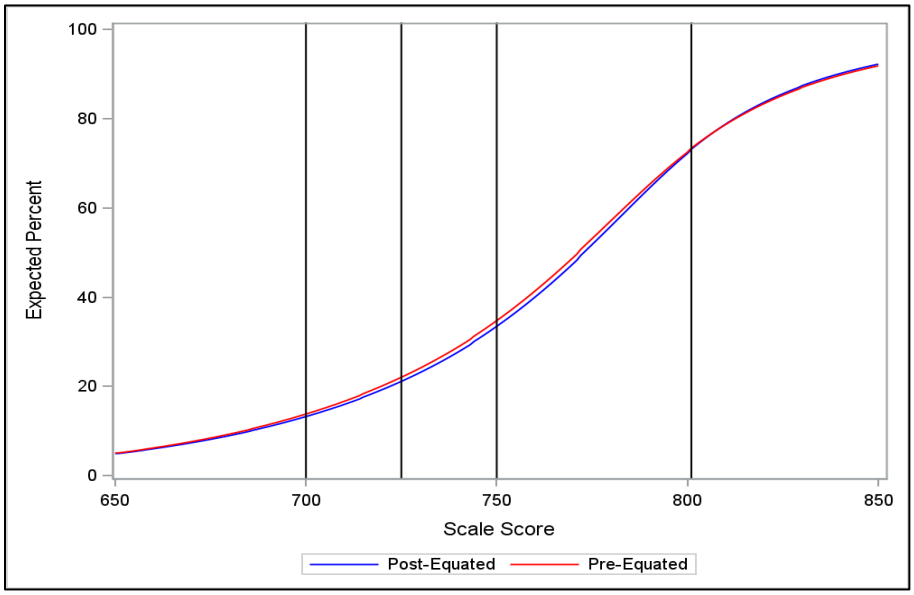


Figure L.6. Post-Equating Check TCCs—Mathematics Grade 8



Appendix M: IRT Item-Level Statistics

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

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	D6	D7
1	XI	GPC	0.464	0.388	0	0.855	0.583	-1.438	–	–	–
2	XI	GPC	0.455	1.709	0	2.715	0.282	-0.534	-2.463	–	–
3	XI	GPC	0.708	0.955	0	1.543	-1.543	–	–	–	–
4	XI	GPC	0.932	0.252	0	0.509	-0.509	–	–	–	–
5	XI	GPC	0.423	-0.109	0	1.089	-1.089	–	–	–	–
6	XI	GPC	0.694	0.156	0	1.332	-1.164	-0.168	–	–	–
7	XI	GPC	0.428	0.690	0	1.361	-0.803	0.395	-0.953	–	–
8	XI	GPC	0.644	0.619	0	2.276	0.798	-0.716	0.402	-2.25	-0.51
9	XI	2PL	0.612	-1.369	–	–	–	–	–	–	–
10	XI	2PL	0.994	-0.978	–	–	–	–	–	–	–
11	XI	2PL	0.893	0.092	–	–	–	–	–	–	–
12	XI	2PL	1.124	0.203	–	–	–	–	–	–	–
13	XI	2PL	0.787	0.145	–	–	–	–	–	–	–
14	XI	2PL	0.471	-0.130	–	–	–	–	–	–	–
15	XI	2PL	0.689	-0.840	–	–	–	–	–	–	–
16	XI	2PL	0.601	1.206	–	–	–	–	–	–	–
17	XI	2PL	0.415	-2.233	–	–	–	–	–	–	–
18	XI	2PL	0.981	0.145	–	–	–	–	–	–	–
19	XI	2PL	0.616	1.283	–	–	–	–	–	–	–
20	XI	2PL	0.628	-0.824	–	–	–	–	–	–	–
21	XI	2PL	1.117	-0.256	–	–	–	–	–	–	–
22	XI	2PL	1.094	0.161	–	–	–	–	–	–	–
23	XI	2PL	1.104	-0.483	–	–	–	–	–	–	–
24	SR	2PL	0.696	-1.703	–	–	–	–	–	–	–
25	SR	2PL	0.914	-1.309	–	–	–	–	–	–	–
26	SR	2PL	0.509	-0.236	–	–	–	–	–	–	–
27	SR	2PL	0.838	-1.325	–	–	–	–	–	–	–
28	SR	2PL	0.629	-0.240	–	–	–	–	–	–	–
29	SR	2PL	0.791	-1.646	–	–	–	–	–	–	–
30	SR	2PL	0.463	0.032	–	–	–	–	–	–	–
31	SR	2PL	0.859	-0.887	–	–	–	–	–	–	–
32	SR	2PL	0.586	-1.999	–	–	–	–	–	–	–
33	SR	2PL	0.880	-0.378	–	–	–	–	–	–	–

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

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	D6	D7
1	CR	GPC	0.519	0.752	0.000	-0.358	0.252	0.107	–	–	–
2	CR	GPC	0.581	1.157	0.000	-0.931	0.068	0.863	–	–	–
3	CR	GPC	0.411	0.534	0.000	-0.268	1.591	-0.873	0.593	-1.213	0.170
4	XI	GPC	0.801	0.573	0.000	-0.218	0.218	–	–	–	–
5	XI	GPC	0.404	-0.335	0.000	0.751	-0.751	–	–	–	–
6	XI	GPC	0.444	0.083	0.000	0.923	-2.323	1.014	0.386	–	–
7	XI	GPC	0.801	-0.188	0.000	0.486	-0.486	–	–	–	–
8	XI	GPC	0.884	0.226	0.000	0.525	-0.525	–	–	–	–
9	XI	GPC	0.544	0.687	0.000	0.973	-0.973	–	–	–	–
10	XI	GPC	0.687	-0.817	0.000	0.426	-0.426	–	–	–	–
11	XI	GPC	0.733	0.550	0.000	0.385	0.442	-0.209	-0.618	–	–
12	XI	2PL	0.970	0.978	–	–	–	–	–	–	–
13	XI	2PL	0.653	-0.372	–	–	–	–	–	–	–
14	XI	2PL	0.868	0.078	–	–	–	–	–	–	–
15	XI	2PL	1.111	0.853	–	–	–	–	–	–	–
16	XI	2PL	1.317	1.064	–	–	–	–	–	–	–
17	XI	2PL	0.533	-0.769	–	–	–	–	–	–	–
18	XI	2PL	0.631	0.314	–	–	–	–	–	–	–
19	XI	2PL	1.267	0.354	–	–	–	–	–	–	–
20	XI	2PL	0.450	0.267	–	–	–	–	–	–	–
21	XI	2PL	0.803	-0.693	–	–	–	–	–	–	–
22	XI	2PL	1.040	0.385	–	–	–	–	–	–	–
23	SR	2PL	0.575	0.294	–	–	–	–	–	–	–
24	SR	2PL	0.779	-0.900	–	–	–	–	–	–	–
25	SR	2PL	0.688	-0.718	–	–	–	–	–	–	–
26	SR	2PL	1.271	-1.046	–	–	–	–	–	–	–
27	SR	2PL	0.693	0.123	–	–	–	–	–	–	–
28	SR	2PL	0.521	0.123	–	–	–	–	–	–	–
29	SR	2PL	0.903	-0.155	–	–	–	–	–	–	–
30	SR	2PL	0.504	1.278	–	–	–	–	–	–	–
31	SR	2PL	0.924	-0.679	–	–	–	–	–	–	–

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

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	D6	D7
1	CR	GPC	0.723	1.373	0	0.934	-0.034	-0.900	–	–	–
2	CR	GPC	0.493	1.032	0	-0.185	0.760	-0.528	-0.047	–	–
3	CR	GPC	0.851	-0.385	0	0.344	-0.344	–	–	–	–
4	XI	GPC	0.580	0.365	0	0.493	-0.493	–	–	–	–
5	XI	GPC	0.580	-0.339	0	0.579	-0.579	–	–	–	–
6	XI	GPC	0.971	0.232	0	0.395	-0.395	–	–	–	–
7	XI	GPC	0.978	-0.056	0	0.373	-0.373	–	–	–	–
8	XI	GPC	0.640	0.842	0	0.122	-0.122	–	–	–	–
9	XI	GPC	0.709	0.794	0	0.742	-0.397	-0.345	–	–	–
10	XI	GPC	0.803	1.135	0	1.249	0.101	-0.197	-1.153	–	–
11	XI	GPC	0.522	1.362	0	-0.624	2.848	-1.030	-0.368	0.034	-0.860
12	XI	2PL	0.796	-0.304	–	–	–	–	–	–	–
13	XI	2PL	0.501	1.021	–	–	–	–	–	–	–
14	XI	2PL	0.676	-1.277	–	–	–	–	–	–	–
15	XI	2PL	0.876	0.233	–	–	–	–	–	–	–
16	XI	2PL	1.064	0.261	–	–	–	–	–	–	–
17	XI	2PL	0.508	-2.591	–	–	–	–	–	–	–
18	XI	2PL	0.925	-0.211	–	–	–	–	–	–	–
19	XI	2PL	0.740	0.580	–	–	–	–	–	–	–
20	XI	2PL	0.627	-0.097	–	–	–	–	–	–	–
21	SR	2PL	0.616	0.767	–	–	–	–	–	–	–
22	SR	2PL	1.049	-0.923	–	–	–	–	–	–	–
23	SR	2PL	0.421	-1.819	–	–	–	–	–	–	–
24	SR	2PL	0.849	-0.812	–	–	–	–	–	–	–
25	SR	2PL	0.524	-1.030	–	–	–	–	–	–	–
26	SR	2PL	0.488	1.259	–	–	–	–	–	–	–
27	SR	2PL	1.054	-0.969	–	–	–	–	–	–	–
28	SR	2PL	0.642	-0.344	–	–	–	–	–	–	–
29	SR	2PL	0.538	0.037	–	–	–	–	–	–	–
30	SR	2PL	0.659	0.150	–	–	–	–	–	–	–

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

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	D6	D7
1	CR	GPC	0.722	-0.077	0.000	0.285	0.667	-0.951	–	–	–
2	CR	GPC	0.596	1.293	0.000	0.021	0.258	0.523	-0.802	–	–
3	CR	GPC	0.437	0.144	0.000	0.087	0.688	0.957	-1.732	–	–
4	CR	GPC	0.575	0.623	0.000	0.172	0.820	-0.991	–	–	–
5	XI	GPC	0.889	0.194	0.000	0.247	-0.247	–	–	–	–
6	XI	GPC	0.863	1.169	0.000	0.638	-0.638	–	–	–	–
7	XI	GPC	0.680	-0.639	0.000	0.658	-0.658	–	–	–	–
8	XI	GPC	0.598	1.190	0.000	0.021	0.752	0.942	0.151	-0.638	-1.228
9	XI	GPC	0.304	0.375	0.000	1.425	-1.425	–	–	–	–
10	XI	GPC	0.478	0.399	0.000	1.487	0.225	-0.720	-0.991	–	–
11	XI	2PL	0.839	-0.746	–	–	–	–	–	–	–
12	XI	2PL	0.704	1.295	–	–	–	–	–	–	–
13	XI	2PL	0.562	-0.882	–	–	–	–	–	–	–
14	XI	2PL	0.762	2.042	–	–	–	–	–	–	–
15	XI	2PL	1.137	0.421	–	–	–	–	–	–	–
16	XI	2PL	0.853	0.760	–	–	–	–	–	–	–
17	XI	2PL	0.501	1.098	–	–	–	–	–	–	–
18	XI	2PL	0.884	0.733	–	–	–	–	–	–	–
19	XI	2PL	0.820	0.723	–	–	–	–	–	–	–
20	XI	2PL	1.034	0.466	–	–	–	–	–	–	–
21	XI	2PL	1.164	0.535	–	–	–	–	–	–	–
22	XI	2PL	1.048	0.953	–	–	–	–	–	–	–
23	XI	2PL	0.908	-0.591	–	–	–	–	–	–	–
24	SR	2PL	0.664	-0.350	–	–	–	–	–	–	–
25	SR	2PL	0.206	1.078	–	–	–	–	–	–	–
26	SR	2PL	0.705	0.669	–	–	–	–	–	–	–
27	SR	2PL	0.501	-0.465	–	–	–	–	–	–	–
28	SR	2PL	0.310	1.621	–	–	–	–	–	–	–
29	SR	2PL	0.364	1.355	–	–	–	–	–	–	–
30	SR	2PL	0.747	-0.812	–	–	–	–	–	–	–
31	SR	2PL	0.863	0.511	–	–	–	–	–	–	–
32	SR	2PL	0.341	1.464	–	–	–	–	–	–	–
33	SR	2PL	0.445	-1.487	–	–	–	–	–	–	–
34	SR	2PL	0.706	-0.892	–	–	–	–	–	–	–
35	SR	2PL	0.741	-1.381	–	–	–	–	–	–	–
36	SR	2PL	0.686	-0.762	–	–	–	–	–	–	–
37	SR	2PL	0.310	1.268	–	–	–	–	–	–	–
38	SR	2PL	0.358	1.873	–	–	–	–	–	–	–
39	SR	2PL	0.889	-1.944	–	–	–	–	–	–	–
40	SR	2PL	0.645	-1.271	–	–	–	–	–	–	–

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

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	D6	D7
1	CR	GPC	0.694	0.826	0	0.156	0.375	-0.531	–	–	–
2	CR	GPC	0.814	1.552	0	1.102	-0.018	-1.084	–	–	–
3	CR	GPC	0.575	0.889	0	0.216	-0.505	0.244	0.046	–	–
4	XI	GPC	0.366	1.963	0	-1.175	1.175	–	–	–	–
5	XI	GPC	0.981	1.184	0	0.210	-0.210	–	–	–	–
6	XI	GPC	0.372	0.681	0	0.657	-0.657	–	–	–	–
7	XI	GPC	1.012	1.598	0	0.487	-0.487	–	–	–	–
8	XI	GPC	0.513	1.159	0	1.552	0.084	-0.600	-1.036	–	–
9	XI	GPC	0.518	1.033	0	1.425	0.216	-0.434	-1.207	–	–
10	XI	GPC	0.598	0.804	0	1.420	-0.054	0.491	0.017	-0.703	-1.171
11	XI	2PL	1.031	-0.114	–	–	–	–	–	–	–
12	XI	2PL	0.750	1.275	–	–	–	–	–	–	–
13	XI	2PL	0.842	-0.047	–	–	–	–	–	–	–
14	XI	2PL	0.964	1.280	–	–	–	–	–	–	–
15	XI	2PL	1.184	1.420	–	–	–	–	–	–	–
16	XI	2PL	0.534	0.838	–	–	–	–	–	–	–
17	XI	2PL	1.450	0.869	–	–	–	–	–	–	–
18	XI	2PL	0.399	0.027	–	–	–	–	–	–	–
19	XI	2PL	1.041	-0.780	–	–	–	–	–	–	–
20	XI	2PL	0.558	0.121	–	–	–	–	–	–	–
21	SR	2PL	0.555	1.625	–	–	–	–	–	–	–
22	SR	2PL	0.454	-0.962	–	–	–	–	–	–	–
23	SR	2PL	0.279	1.924	–	–	–	–	–	–	–
24	SR	2PL	0.654	1.069	–	–	–	–	–	–	–
25	SR	2PL	0.826	0.791	–	–	–	–	–	–	–
26	SR	2PL	0.252	0.674	–	–	–	–	–	–	–
27	SR	2PL	0.596	-0.242	–	–	–	–	–	–	–
28	SR	2PL	0.366	1.908	–	–	–	–	–	–	–
29	SR	2PL	0.669	-0.652	–	–	–	–	–	–	–

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

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	D6	D7
1	CR	GPC	0.668	1.506	0.000	-0.474	1.130	-0.656	–	–	–
2	CR	GPC	0.694	2.260	0.000	0.382	0.487	0.306	-1.174	–	–
3	CR	GPC	0.537	1.435	0.000	-0.119	-0.641	0.759	–	–	–
4	XI	GPC	0.392	0.297	0.000	0.973	-0.973	–	–	–	–
5	XI	GPC	0.818	1.503	0.000	0.809	-0.809	–	–	–	–
6	XI	GPC	0.318	2.800	0.000	1.720	-1.720	–	–	–	–
7	XI	GPC	0.397	0.754	0.000	1.167	-0.541	-0.149	-0.477	–	–
8	XI	GPC	0.717	1.605	0.000	1.717	0.057	-0.243	-1.531	–	–
9	XI	GPC	0.559	0.837	0.000	0.362	-0.362	–	–	–	–
10	XI	GPC	0.494	0.311	0.000	0.728	-0.728	–	–	–	–
11	XI	GPC	0.595	1.690	0.000	1.012	0.658	-0.336	-0.042	-0.260	-1.031
12	XI	GPC	0.391	1.415	0.000	2.126	-2.126	–	–	–	–
13	XI	GPC	0.289	2.139	0.000	1.999	-1.999	–	–	–	–
14	XI	GPC	0.489	1.483	0.000	0.642	-0.642	–	–	–	–
15	XI	GPC	0.683	1.522	0.000	0.491	-0.491	–	–	–	–
16	XI	2PL	1.336	1.477	–	–	–	–	–	–	–
17	XI	2PL	0.847	0.432	–	–	–	–	–	–	–
18	XI	2PL	1.011	1.172	–	–	–	–	–	–	–
19	XI	2PL	1.202	1.799	–	–	–	–	–	–	–
20	XI	2PL	0.674	1.114	–	–	–	–	–	–	–
21	XI	2PL	0.629	1.024	–	–	–	–	–	–	–
22	XI	2PL	0.907	0.967	–	–	–	–	–	–	–
23	XI	2PL	0.842	2.711	–	–	–	–	–	–	–
24	SR	2PL	0.352	0.573	–	–	–	–	–	–	–
25	SR	2PL	0.321	0.916	–	–	–	–	–	–	–
26	SR	2PL	0.529	1.042	–	–	–	–	–	–	–
27	SR	2PL	0.495	1.344	–	–	–	–	–	–	–
28	SR	2PL	0.680	-1.703	–	–	–	–	–	–	–
29	SR	2PL	0.633	-1.005	–	–	–	–	–	–	–
30	SR	2PL	0.518	1.671	–	–	–	–	–	–	–
31	SR	2PL	0.718	-0.651	–	–	–	–	–	–	–
32	SR	2PL	0.358	-0.129	–	–	–	–	–	–	–
33	SR	2PL	0.314	-0.184	–	–	–	–	–	–	–
34	SR	2PL	0.933	-1.160	–	–	–	–	–	–	–
35	SR	2PL	0.557	-1.532	–	–	–	–	–	–	–
36	SR	2PL	0.705	-0.613	–	–	–	–	–	–	–

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

Item	Item Type	Model	A	B	D1	D2	D3	D4	Misfit Flag
PCR 1 WE	CR	GPC	0.721	2.002	0	2.269	-0.285	-1.984	No
PCR 1 WKL	CR	GPC	0.609	2.226	0	1.944	0.271	-2.215	No
PCR 2 WE	CR	GPC	0.848	2.084	0	1.883	-0.185	-1.698	No
PCR 2 WKL	CR	GPC	0.711	2.231	0	2.036	-0.509	-1.527	No
5	XI	GPC	0.307	1.914	0	-0.703	0.703	–	Yes
6	XI	GPC	0.267	1.073	0	-0.515	0.515	–	Yes
7	XI	GPC	0.217	1.901	0	-0.079	0.079	–	No
8	XI	GPC	0.491	-0.190	0	0.829	-0.829	–	No
9	XI	GPC	0.552	0.362	0	0.889	-0.889	–	No
10	XI	GPC	0.544	-1.281	0	-2.770	2.770	–	No
11	XI	GPC	0.516	-0.551	0	-1.778	1.778	–	Yes
12	XI	GPC	0.518	-0.722	0	0.098	-0.098	–	Yes
13	XI	GPC	0.391	0.486	0	-0.797	0.797	–	No
14	XI	GPC	0.894	-1.181	0	0.350	-0.350	–	Yes
15	XI	GPC	0.617	-1.596	0	0.508	-0.508	–	Yes
16	XI	GPC	0.447	0.951	0	-0.877	0.877	–	No
17	XI	GPC	0.470	-0.244	0	-0.927	0.927	–	Yes
18	XI	GPC	0.366	-0.392	0	-1.584	1.584	–	Yes
19	XI	GPC	0.392	-0.328	0	-1.627	1.627	–	Yes
20	XI	GPC	0.400	0.301	0	-1.532	1.532	–	Yes
21	XI	GPC	0.545	0.204	0	0.898	-0.898	–	No
22	XI	GPC	0.725	0.028	0	1.287	-1.287	–	No
23	XI	GPC	0.502	1.329	0	1.007	-1.007	–	No

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

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
PCR 1 WE	CR	GPC	0.752	0.907	0	1.395	-0.170	-1.225	–	No
PCR 1 WKL	CR	GPC	0.742	1.131	0	1.292	-0.092	-1.200	–	No
PCR 2 WE	CR	GPC	0.796	1.889	0	2.259	-0.070	-0.945	-1.244	No
PCR 2 WKL	CR	GPC	0.759	1.505	0	1.429	-0.298	-1.132	–	No
5	XI	GPC	0.309	-0.782	0	-1.064	1.064	–	–	No
6	XI	GPC	0.237	1.304	0	0.509	-0.509	–	–	Yes
7	XI	GPC	0.413	0.067	0	-0.654	0.654	–	–	No
8	XI	GPC	0.400	-0.301	0	-0.052	0.052	–	–	No
9	XI	GPC	0.523	0.257	0	1.486	-1.486	–	–	No
10	XI	GPC	0.528	-1.272	0	-1.734	1.734	–	–	Yes
11	XI	GPC	0.239	1.632	0	-0.547	0.547	–	–	No
12	XI	GPC	0.976	-0.506	0	0.753	-0.753	–	–	No
13	XI	GPC	0.609	-0.065	0	0.141	-0.141	–	–	No
14	XI	GPC	0.201	2.052	0	-3.226	3.226	–	–	Yes
15	XI	GPC	0.662	0.199	0	0.508	-0.508	–	–	No
16	XI	GPC	0.373	0.224	0	0.492	-0.492	–	–	No
17	XI	GPC	0.598	-1.241	0	0.424	-0.424	–	–	Yes
18	XI	GPC	0.532	0.248	0	-2.767	2.767	–	–	No
19	XI	GPC	0.294	1.408	0	-0.292	0.292	–	–	Yes
20	XI	GPC	0.510	0.142	0	-1.549	1.549	–	–	Yes
21	XI	GPC	0.544	-0.458	0	0.147	-0.147	–	–	No
22	XI	GPC	0.347	-0.752	0	-0.321	0.321	–	–	Yes
23	XI	GPC	0.337	-1.246	0	-3.278	3.278	–	–	Yes
24	XI	GPC	0.788	0.024	0	0.067	-0.067	–	–	Yes
25	XI	GPC	0.874	-0.793	0	0.730	-0.730	–	–	No

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

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
PCR 1 WE	CR	GPC	0.806	1.452	0	0.983	0.006	-0.988	–	No
PCR 1 WKL	CR	GPC	0.758	1.604	0	0.913	0.013	-0.926	–	No
PCR 2 WE	CR	GPC	0.811	2.335	0	1.971	0.009	-0.707	-1.273	Yes
PCR 2 WKL	CR	GPC	0.806	1.927	0	1.551	-0.392	-1.159	–	No
5	XI	GPC	0.590	-0.843	0	-1.360	1.360	–	–	Yes
6	XI	GPC	0.573	-0.175	0	-0.405	0.405	–	–	Yes
7	XI	GPC	0.432	0.304	0	-0.962	0.962	–	–	Yes
8	XI	GPC	0.415	0.359	0	0.000	0.000	–	–	No
9	XI	GPC	0.263	1.445	0	-0.375	0.375	–	–	No
10	XI	GPC	0.346	1.602	0	-1.446	1.446	–	–	Yes
11	XI	GPC	0.910	-0.302	0	-0.033	0.033	–	–	Yes
12	XI	GPC	0.335	0.776	0	-1.398	1.398	–	–	Yes
13	XI	GPC	0.580	1.322	0	0.671	-0.671	–	–	No
14	XI	GPC	0.258	0.899	0	-2.679	2.679	–	–	Yes
15	XI	GPC	0.270	1.440	0	-2.614	2.614	–	–	Yes
16	XI	GPC	0.377	1.205	0	-0.676	0.676	–	–	Yes
17	XI	GPC	0.256	1.515	0	1.285	-1.285	–	–	Yes
18	XI	GPC	0.191	-0.078	0	-2.488	2.488	–	–	No
19	XI	GPC	0.454	0.250	0	2.001	-2.001	–	–	Yes
20	XI	GPC	0.263	2.301	0	-1.673	1.673	–	–	No
21	XI	GPC	0.618	0.280	0	0.997	-0.997	–	–	Yes
22	XI	GPC	0.403	0.500	0	0.784	-0.784	–	–	No
23	XI	GPC	0.372	0.996	0	-0.075	0.075	–	–	Yes
24	XI	GPC	0.499	0.476	0	1.077	-1.077	–	–	No

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

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
PCR 1 WE	CR	GPC	0.923	1.893	0	1.768	0.389	-0.707	-1.450	No
PCR 1 WKL	CR	GPC	0.857	1.359	0	1.230	-0.037	-1.193	–	No
PCR 2 WE	CR	GPC	0.859	1.676	0	1.671	0.404	-0.753	-1.322	No
PCR 2 WKL	CR	GPC	0.842	1.300	0	1.052	0.013	-1.065	–	No
5	XI	GPC	0.270	1.380	0	-1.536	1.536	–	–	No
6	XI	GPC	0.311	0.328	0	-2.174	2.174	–	–	Yes
7	XI	GPC	0.360	0.985	0	-0.777	0.777	–	–	Yes
8	XI	GPC	0.552	-1.093	0	0.278	-0.278	–	–	No
9	XI	GPC	0.275	1.578	0	1.362	-1.362	–	–	Yes
10	XI	GPC	0.549	-0.556	0	-1.314	1.314	–	–	Yes
11	XI	GPC	0.332	1.129	0	-0.112	0.112	–	–	No
12	XI	GPC	0.245	-0.388	0	-4.513	4.513	–	–	Yes
13	XI	GPC	0.330	1.227	0	1.167	-1.167	–	–	Yes
14	XI	GPC	0.414	-0.087	0	1.519	-1.519	–	–	No
15	XI	GPC	0.584	-0.097	0	0.799	-0.799	–	–	Yes
16	XI	GPC	0.326	0.599	0	0.385	-0.385	–	–	No
17	XI	GPC	0.518	1.153	0	-0.084	0.084	–	–	No
18	XI	GPC	0.423	-0.648	0	-1.267	1.267	–	–	Yes
19	XI	GPC	0.454	0.531	0	-1.247	1.247	–	–	Yes
20	XI	GPC	0.625	0.190	0	0.219	-0.219	–	–	Yes
21	XI	GPC	0.421	0.323	0	-1.377	1.377	–	–	No
22	XI	GPC	0.566	0.245	0	-0.759	0.759	–	–	Yes
23	XI	GPC	0.360	-0.768	0	-0.603	0.603	–	–	Yes
24	XI	GPC	0.233	-0.240	0	2.072	-2.072	–	–	No
25	XI	GPC	0.263	0.264	0	-2.613	2.613	–	–	No
26	XI	GPC	0.201	0.729	0	-1.898	1.898	–	–	Yes

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

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
PCR 1 WE	CR	GPC	1.072	1.602	0	1.713	0.298	-0.547	-1.465	Yes
PCR 1 WKL	CR	GPC	1.045	1.108	0	1.353	-0.253	-1.099	–	No
PCR 2 WE	CR	GPC	0.726	1.087	0	1.368	0.648	-0.454	-1.561	No
PCR 2 WKL	CR	GPC	0.762	0.657	0	0.744	0.143	-0.887	–	No
5	XI	GPC	0.435	-0.410	0	-0.496	0.496	–	–	Yes
6	XI	GPC	0.312	1.201	0	-0.340	0.340	–	–	Yes
7	XI	GPC	0.450	0.831	0	-0.577	0.577	–	–	Yes
8	XI	GPC	0.609	0.507	0	1.405	-1.405	–	–	No
9	XI	GPC	0.467	0.283	0	-0.817	0.817	–	–	Yes
10	XI	GPC	0.452	0.680	0	2.243	-2.243	–	–	No
11	XI	GPC	0.786	-0.078	0	0.590	-0.590	–	–	No
12	XI	GPC	0.615	-0.662	0	-1.056	1.056	–	–	Yes
13	XI	GPC	0.300	0.948	0	-0.764	0.764	–	–	No
14	XI	GPC	0.728	0.275	0	1.187	-1.187	–	–	No
15	XI	GPC	0.281	0.374	0	-0.579	0.579	–	–	No
16	XI	GPC	0.367	-0.363	0	-0.668	0.668	–	–	No
17	XI	GPC	0.430	0.648	0	0.319	-0.319	–	–	Yes
18	XI	GPC	0.434	0.756	0	-1.171	1.171	–	–	Yes
19	XI	GPC	0.654	1.284	0	0.646	-0.646	–	–	No
20	XI	GPC	0.560	0.341	0	0.324	-0.324	–	–	No
21	XI	GPC	0.251	1.574	0	-4.010	4.010	–	–	Yes
22	XI	GPC	0.384	1.455	0	-0.380	0.380	–	–	No
23	XI	GPC	0.295	-0.222	0	-1.947	1.947	–	–	No
24	XI	GPC	0.356	0.373	0	0.600	-0.600	–	–	No
25	XI	GPC	0.565	-0.266	0	-0.471	0.471	–	–	Yes
26	XI	GPC	0.286	0.879	0	-1.936	1.936	–	–	Yes
27	XI	GPC	0.518	0.952	0	0.671	-0.671	–	–	No

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

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
PCR 1 WE	CR	GPC	0.859	0.917	0	1.464	0.865	-0.618	-1.710	Yes
PCR 1 WKL	CR	GPC	0.875	0.352	0	0.910	0.287	-1.197	–	Yes
PCR 2 WE	CR	GPC	0.780	0.927	0	1.395	0.738	-0.551	-1.582	No
PCR 2 WKL	CR	GPC	0.769	0.566	0	0.612	0.290	-0.902	–	Yes
5	XI	GPC	0.973	-0.244	0	0.788	-0.788	–	–	No
6	XI	GPC	0.489	-0.064	0	-0.307	0.307	–	–	No
7	XI	GPC	0.412	0.231	0	0.488	-0.488	–	–	Yes
8	XI	GPC	0.456	0.490	0	1.413	-1.413	–	–	No
9	XI	GPC	0.355	0.703	0	-1.934	1.934	–	–	No
10	XI	GPC	0.277	1.328	0	-1.850	1.850	–	–	Yes
11	XI	GPC	0.474	0.406	0	0.830	-0.830	–	–	No
12	XI	GPC	0.515	-1.042	0	1.676	-1.676	–	–	No
13	XI	GPC	0.480	0.086	0	-1.072	1.072	–	–	Yes
14	XI	GPC	0.277	1.070	0	-3.764	3.764	–	–	No
15	XI	GPC	0.394	-0.276	0	-0.237	0.237	–	–	No
16	XI	GPC	0.290	2.754	0	2.576	-2.576	–	–	No
17	XI	GPC	0.330	0.293	0	-3.241	3.241	–	–	Yes
18	XI	GPC	0.428	-1.118	0	1.308	-1.308	–	–	No
19	XI	GPC	0.240	0.317	0	-2.066	2.066	–	–	Yes
20	XI	GPC	0.224	1.128	0	-7.027	7.027	–	–	Yes
21	XI	GPC	0.357	0.766	0	-3.756	3.756	–	–	Yes
22	XI	GPC	0.632	-0.430	0	1.378	-1.378	–	–	No
23	XI	GPC	0.221	2.188	0	-0.682	0.682	–	–	Yes
24	XI	GPC	0.430	-0.094	0	-1.488	1.488	–	–	Yes
25	XI	GPC	0.318	-0.354	0	-0.434	0.434	–	–	Yes
26	XI	GPC	0.586	-0.639	0	-1.451	1.451	–	–	No
27	XI	GPC	0.285	1.142	0	-4.795	4.795	–	–	Yes

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

Item	Item Type	Model	B	D1	D2	D3	D4	Infit	Outfit
1	CR	Rasch	0.019	0	-0.718	0.408	0.310	1.05	0.99
2	CR	Rasch	0.218	0	-0.168	-0.452	0.620	0.92	0.92
3	CR	Rasch	0.054	0	-0.904	0.694	0.210	1.05	1.02
4	CR	Rasch	1.006	0	-1.416	0.402	1.014	0.83	0.79
5	XI	Rasch	-0.425	0	0.234	-0.234	–	0.89	0.86
6	XI	Rasch	-0.170	0	0.489	-0.489	–	0.88	0.91
7	XI	Rasch	-0.364	0	0.173	-0.173	–	0.86	0.85
8	XI	Rasch	-0.387	0	0.756	-0.756	–	0.93	0.92
9	XI	Rasch	0.033	0	0.541	-0.541	–	1.19	1.32
10	XI	Rasch	0.221	0	-0.996	0.996	–	0.91	0.89
11	XI	Rasch	-0.479	0	0.577	-0.577	–	0.88	0.83
12	XI	Rasch	-0.182	0	0.988	-0.988	–	0.96	0.99
13	XI	Rasch	-0.321	0	0.380	-0.380	–	1.27	1.47
14	XI	Rasch	-0.192	0	-1.290	1.290	–	0.87	0.87
15	XI	Rasch	-0.290	0	0.577	-0.577	–	1.18	1.37
16	XI	Rasch	-0.755	0	-0.773	0.773	–	0.94	0.94
17	XI	Rasch	-0.653	0	0.156	-0.156	–	0.96	0.97
18	XI	Rasch	-0.287	0	0.339	-0.339	–	0.98	0.99
19	XI	Rasch	-0.295	0	0.767	-0.767	–	0.92	0.86
20	XI	Rasch	0.119	0	0.204	-0.204	–	0.99	0.99
21	XI	Rasch	-0.081	0	1.273	-1.273	–	1.13	1.34
22	XI	Rasch	0.272	0	0.121	-0.121	–	0.98	1.08
23	XI	Rasch	0.953	0	-0.148	0.148	–	1.15	1.45
24	XI	Rasch	0.953	0	-0.148	0.148	–	1.15	1.45

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

Item	Item Type	Model	B	D1	D2	D3	D4	D5	Infit	Outfit
1	CR	Rasch	-0.102	0	-0.741	0.345	0.396	–	0.91	0.88
2	CR	Rasch	0.237	0	-0.712	-0.910	-0.006	1.628	0.65	0.65
3	CR	Rasch	-0.605	0	-0.812	0.608	0.204	–	0.93	0.92
4	CR	Rasch	0.069	0	0.133	-0.881	0.748	–	0.93	0.88
5	XI	Rasch	-0.295	0	1.031	-1.031	–	–	0.95	1.00
6	XI	Rasch	-0.400	0	0.367	-0.367	–	–	0.79	0.75
7	XI	Rasch	-0.347	0	-0.715	0.715	–	–	0.91	0.91
8	XI	Rasch	-0.406	0	0.735	-0.735	–	–	0.94	0.93
9	XI	Rasch	0.125	0	1.231	-1.231	–	–	0.98	1.00
10	XI	Rasch	0.770	0	0.275	-0.275	–	–	1.30	1.74
11	XI	Rasch	-1.203	0	-0.136	0.136	–	–	0.83	0.8
12	XI	Rasch	-0.496	0	-0.650	0.650	–	–	1.07	1.07
13	XI	Rasch	-0.162	0	0.908	-0.908	–	–	0.99	0.99
14	XI	Rasch	-1.167	0	0.668	-0.668	–	–	1.03	1.06
15	XI	Rasch	1.334	0	-0.927	0.927	–	–	1.12	1.28
16	XI	Rasch	0.078	0	1.250	-1.250	–	–	1.18	1.29
17	XI	Rasch	0.576	0	0.136	-0.136	–	–	1.06	1.25
18	XI	Rasch	-0.359	0	0.166	-0.166	–	–	0.99	1.01
19	XI	Rasch	0.073	0	0.771	-0.771	–	–	0.97	0.98
20	XI	Rasch	-0.538	0	0.784	-0.784	–	–	0.96	0.94
21	XI	Rasch	0.110	0	0.404	-0.404	–	–	0.99	0.97
22	XI	Rasch	-0.626	0	0.399	-0.399	–	–	0.94	0.94
23	XI	Rasch	0.090	0	0.704	-0.704	–	–	1.10	1.11
24	XI	Rasch	1.102	0	-0.997	0.997	–	–	1.03	1.05
25	XI	Rasch	0.261	0	-0.260	0.260	–	–	1.11	1.16
26	XI	Rasch	0.261	0	-0.260	0.260	–	–	1.11	1.16

Appendix N: TCC, TIC, and CSEM Curves

Figure N.1. Mathematics Grade 3 TCC

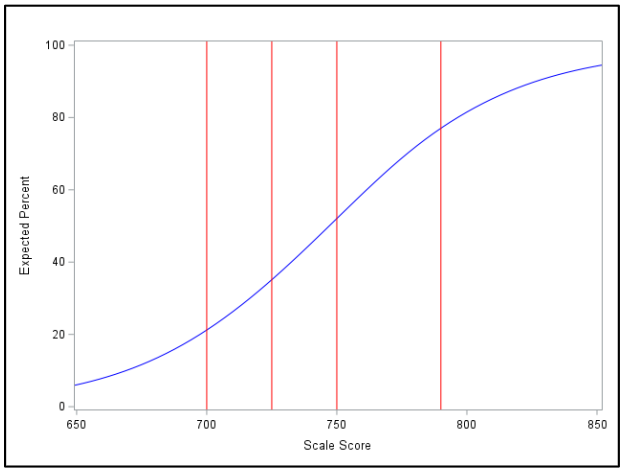


Figure N.2. Mathematics Grade 3 TIC

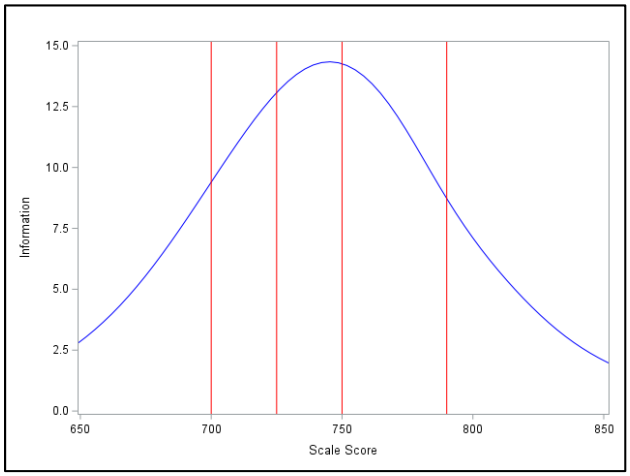


Figure N.3. Mathematics Grade 3 CSEM Curve

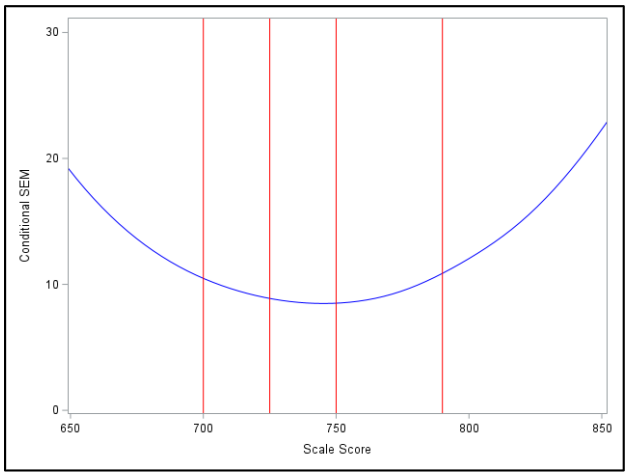


Figure N.4. Mathematics Grade 4 TCC

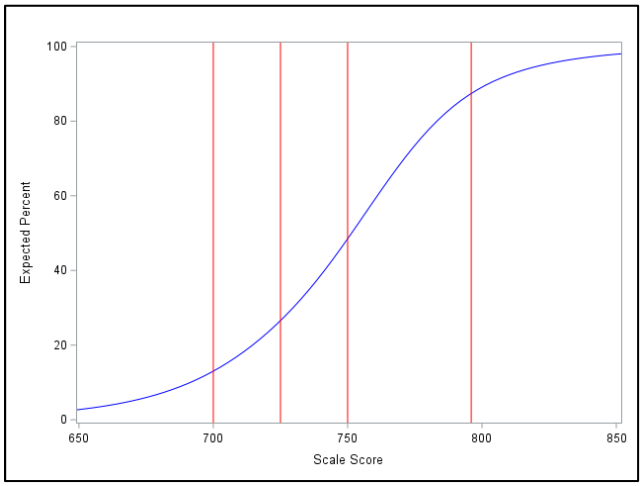


Figure N.5. Mathematics Grade 4 TIC

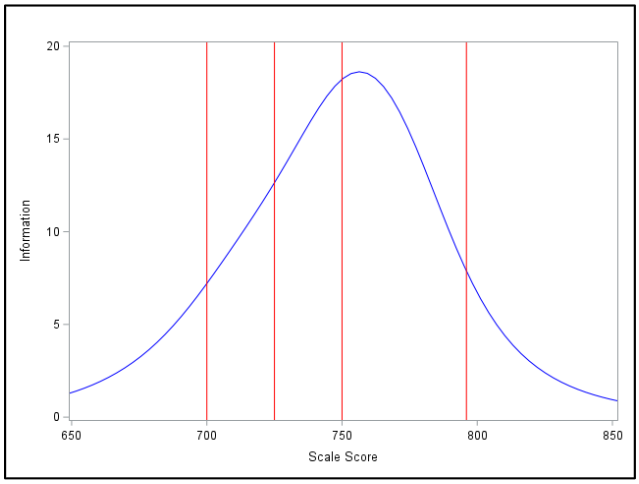


Figure N.6. Mathematics Grade 4 CSEM Curve

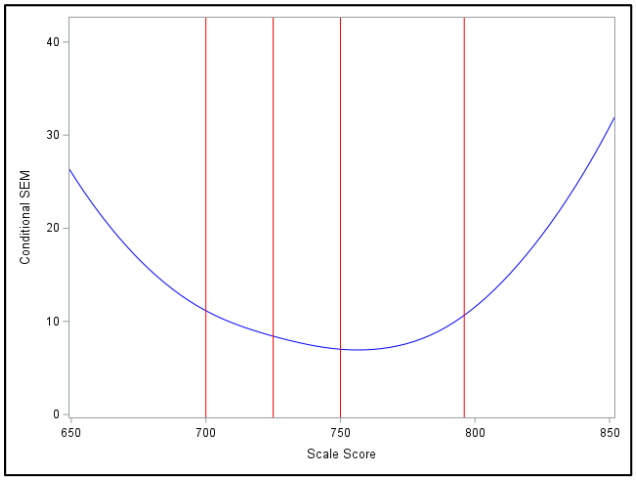


Figure N.7. Mathematics Grade 5 TCC

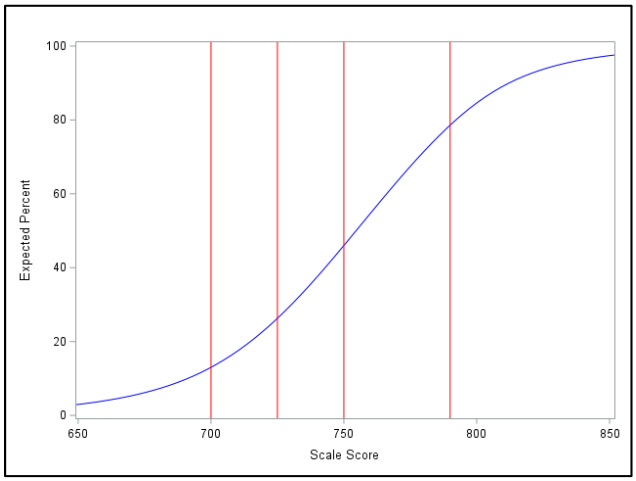


Figure N.8. Mathematics Grade 5 TIC

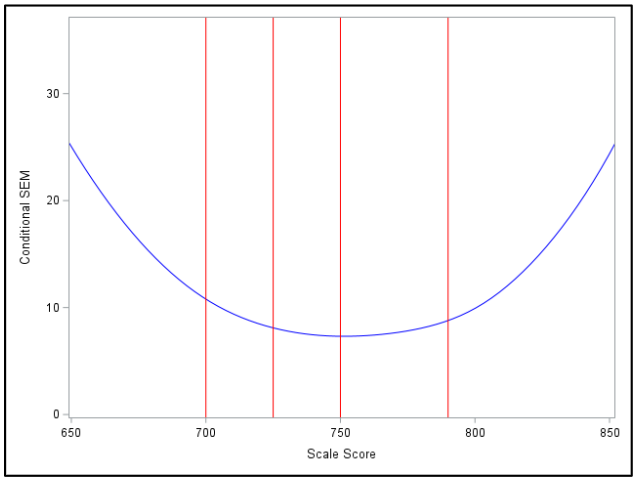


Figure N.9. Mathematics Grade 5 CSEM Curve

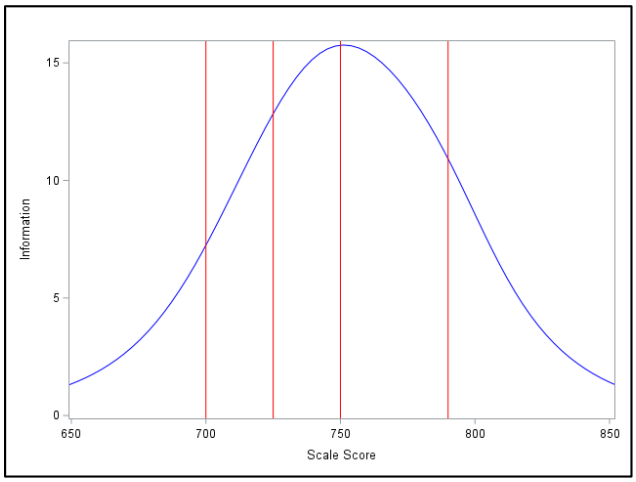


Figure N.10. Mathematics Grade 6 TCC

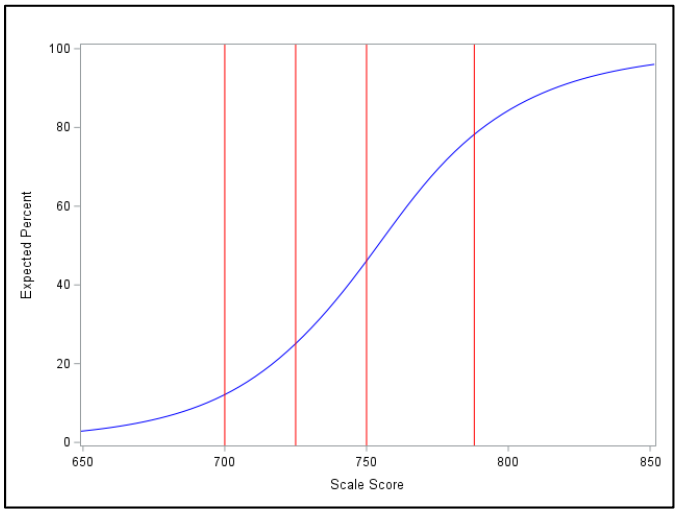


Figure N.11. Mathematics Grade 6 TIC

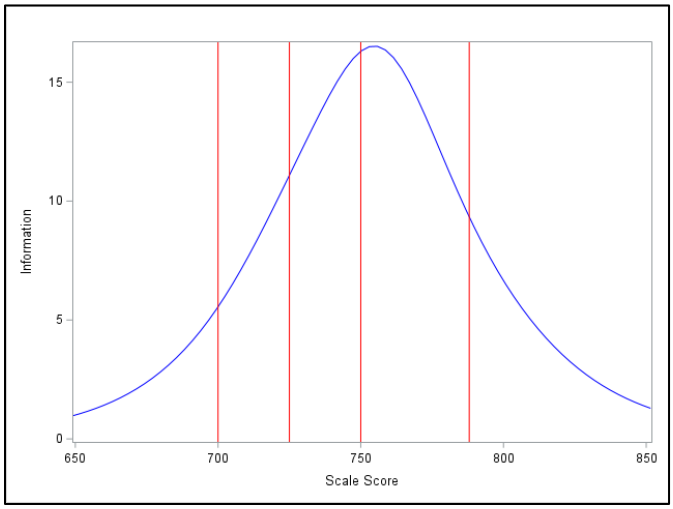


Figure N.12. Mathematics Grade 6 CSEM Curve

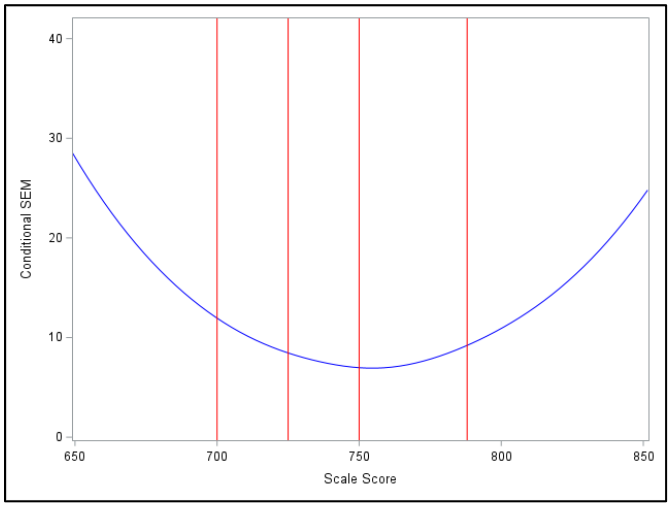


Figure N.13. Mathematics Grade 7 TCC

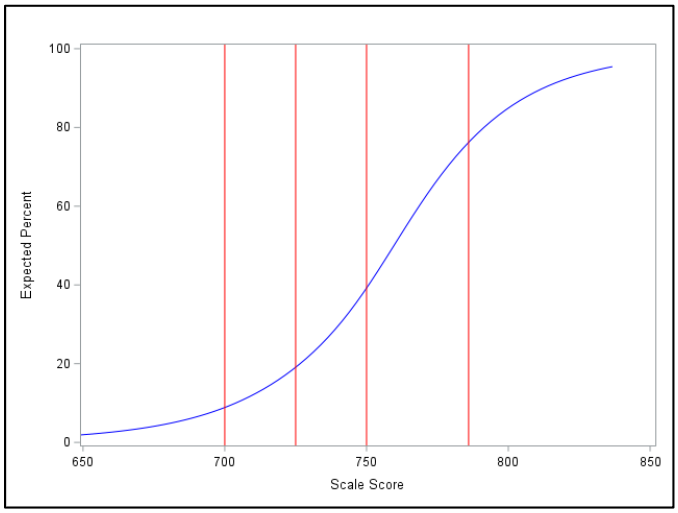


Figure N.14. Mathematics Grade 7 TIC

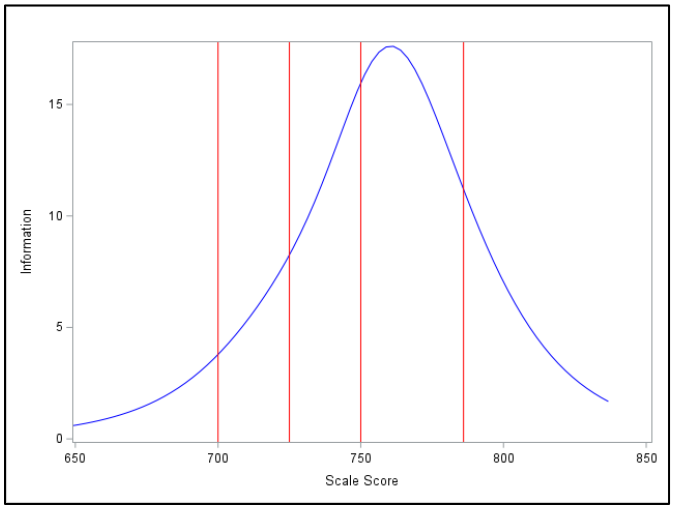


Figure N.15. Mathematics Grade 7 CSEM Curve

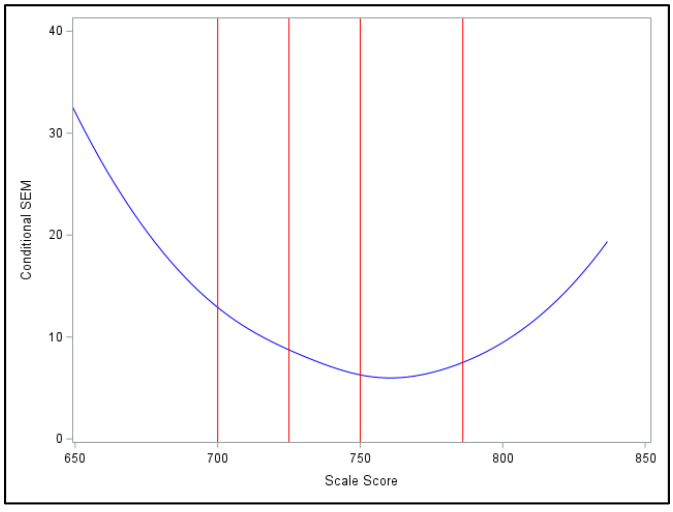


Figure N.16. Mathematics Grade 8 TCC

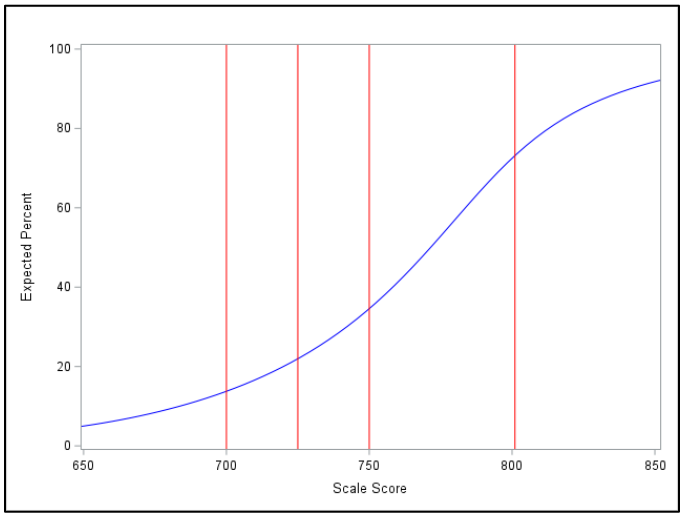


Figure N.17. Mathematics Grade 8 TIC

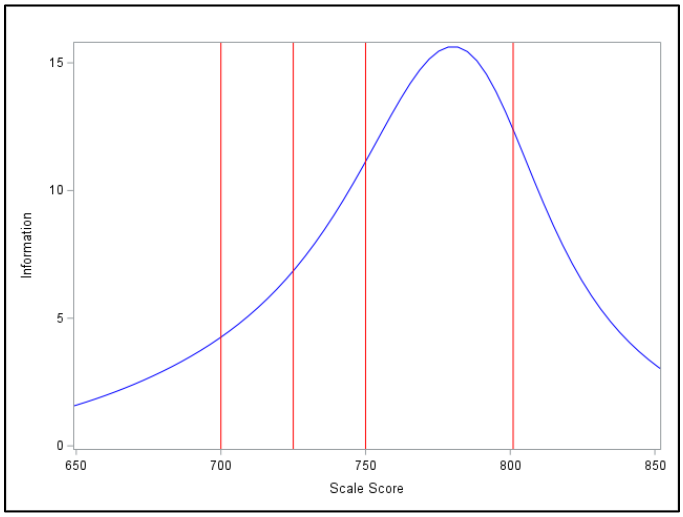


Figure N.18. Mathematics Grade 8 CSEM Curve

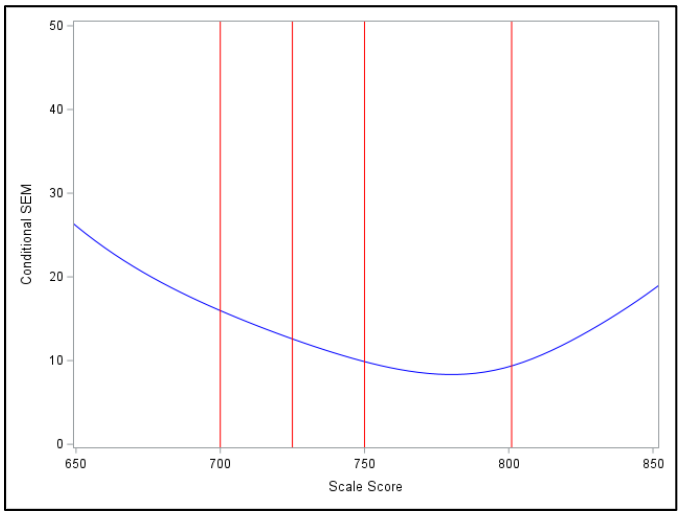


Figure N.19. ELA Grade 3 TCC

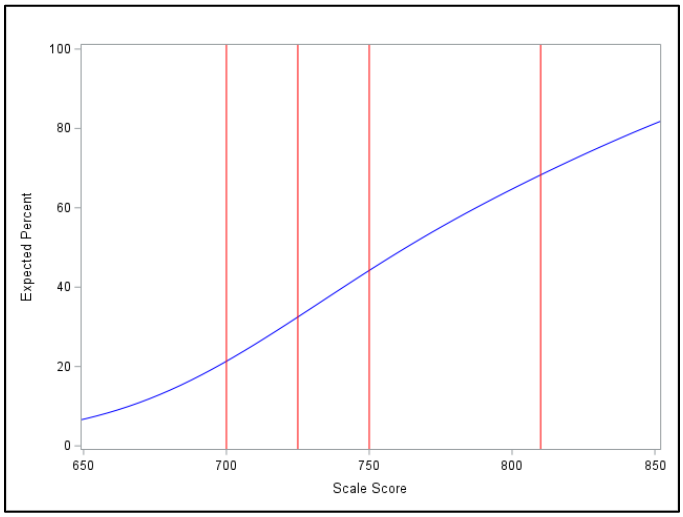


Figure N.20. ELA Grade 3 TIC

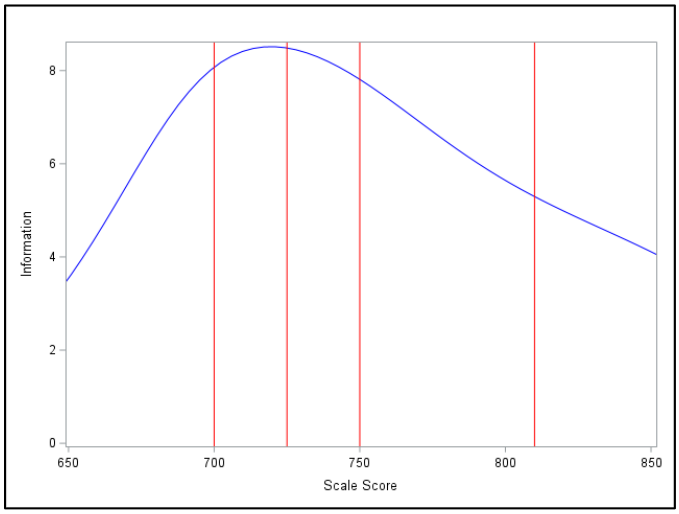


Figure N.21. ELA Grade 3 CSEM Curve

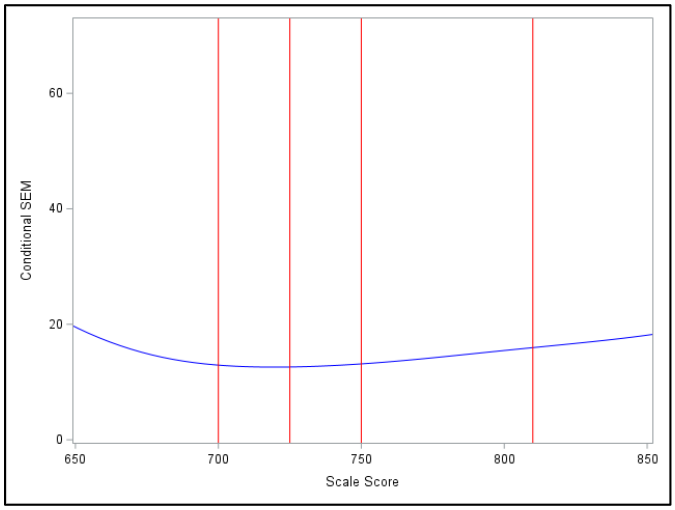


Figure N.22. ELA Grade 4 TCC

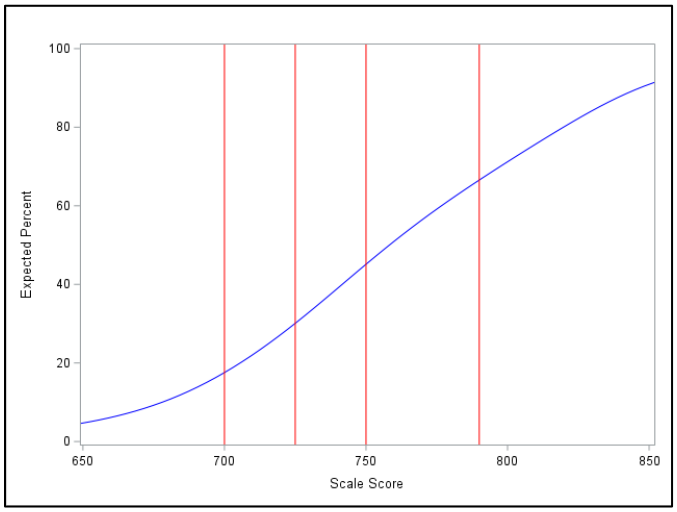


Figure N.23. ELA Grade 4 TIC

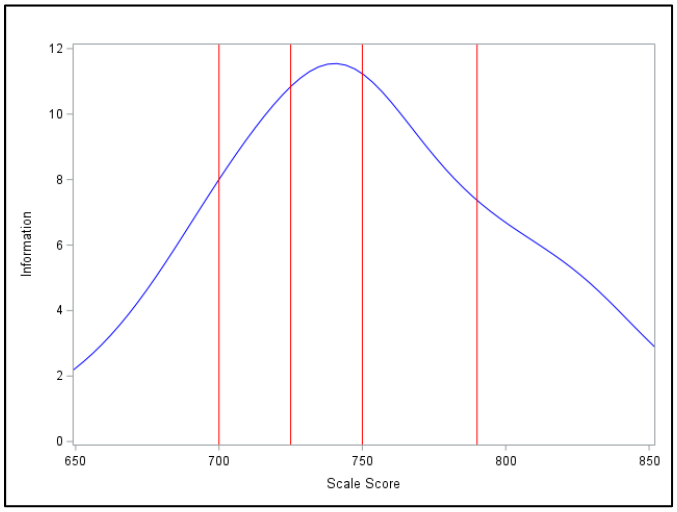


Figure N.24. ELA Grade 4 CSEM Curve

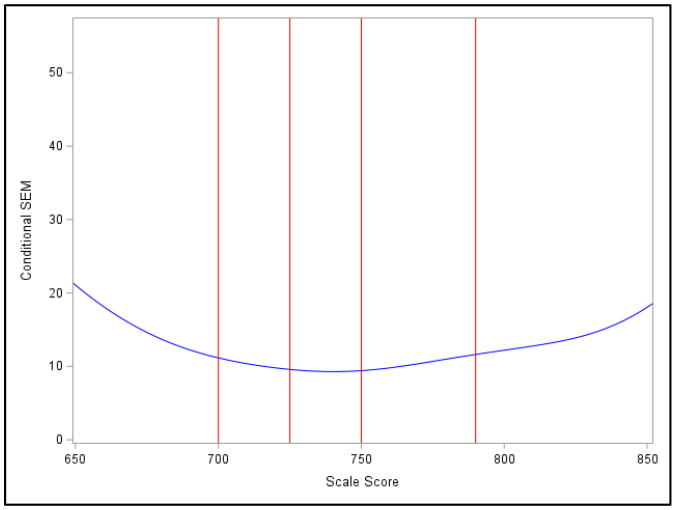


Figure N.25. ELA Grade 5 TCC

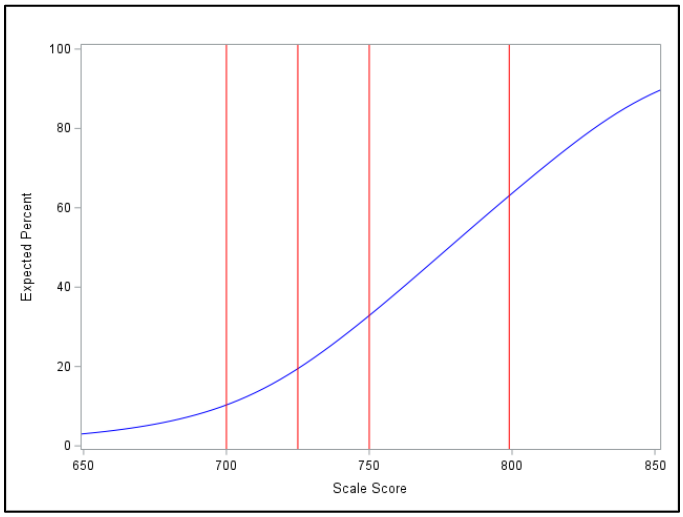


Figure N.26. ELA Grade 5 TIC

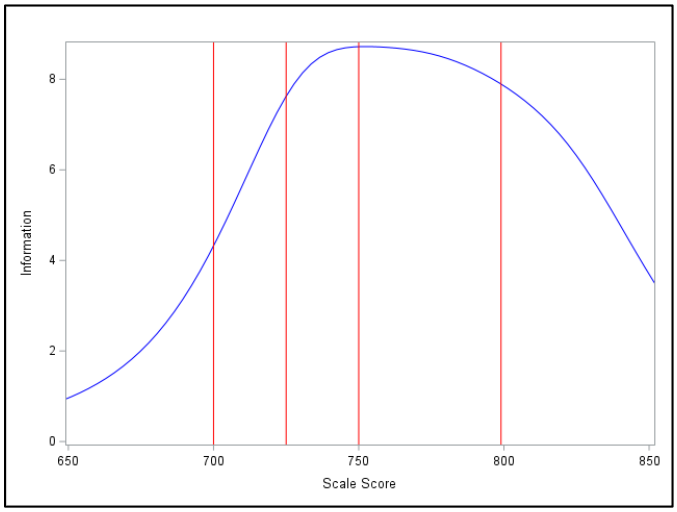


Figure N.27. ELA Grade 5 CSEM Curve

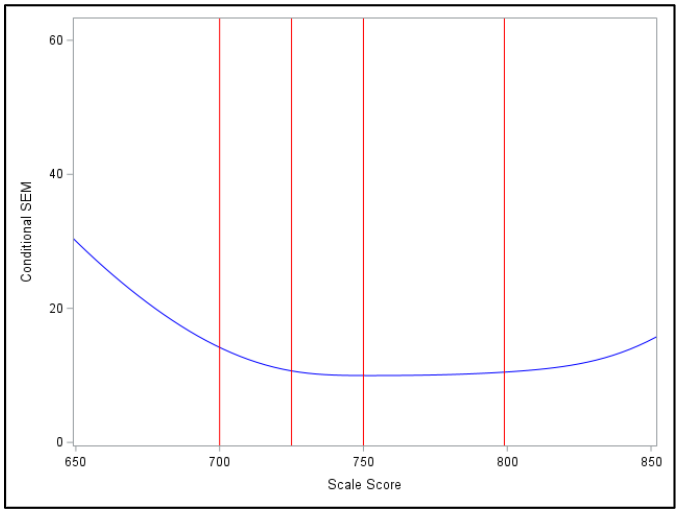


Figure N.28. ELA Grade 6 TCC

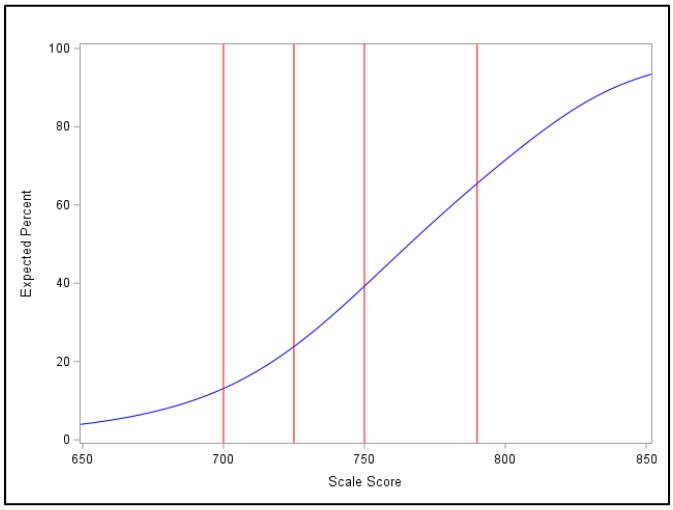


Figure N.29. ELA Grade 6 TIC

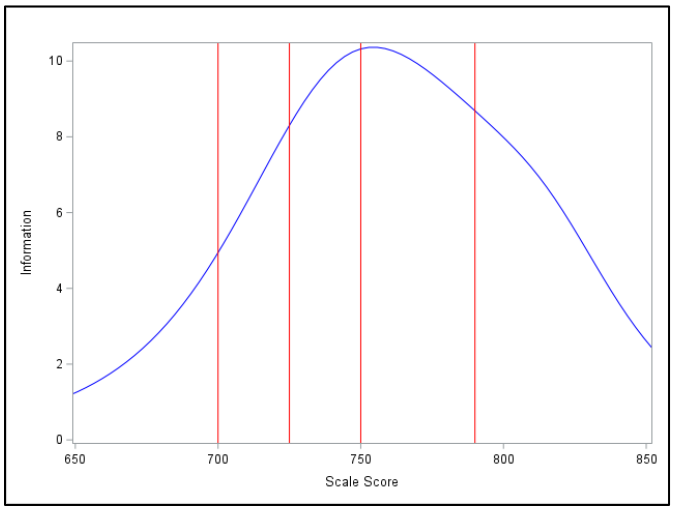


Figure N.30. ELA Grade 6 CSEM Curve

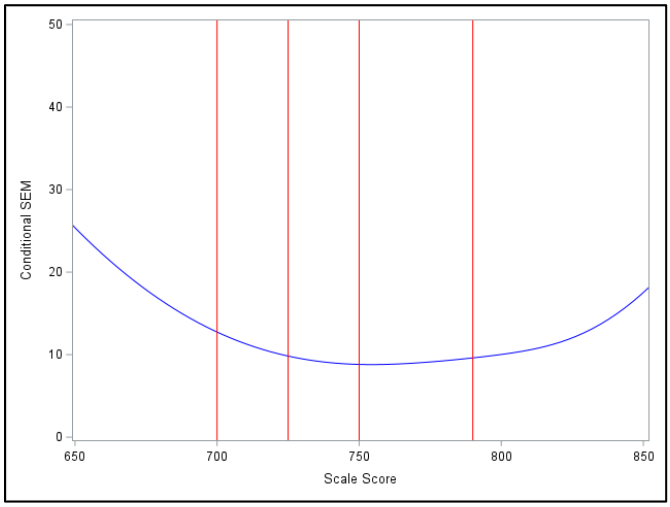


Figure N.31. ELA Grade 7 TCC

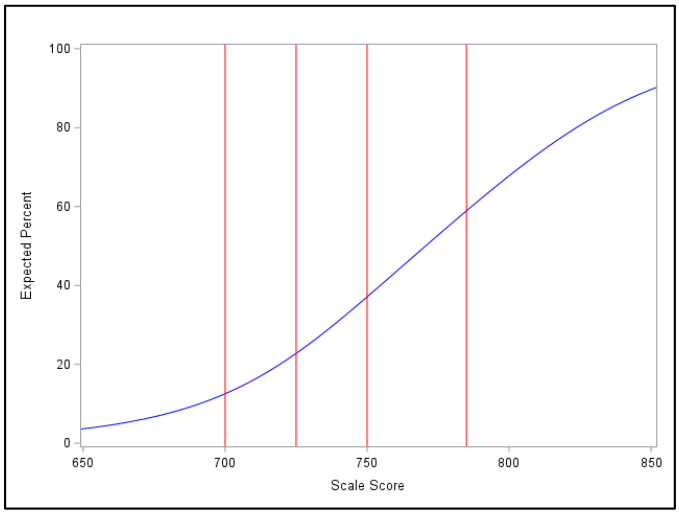


Figure N.32. ELA Grade 7 TIC

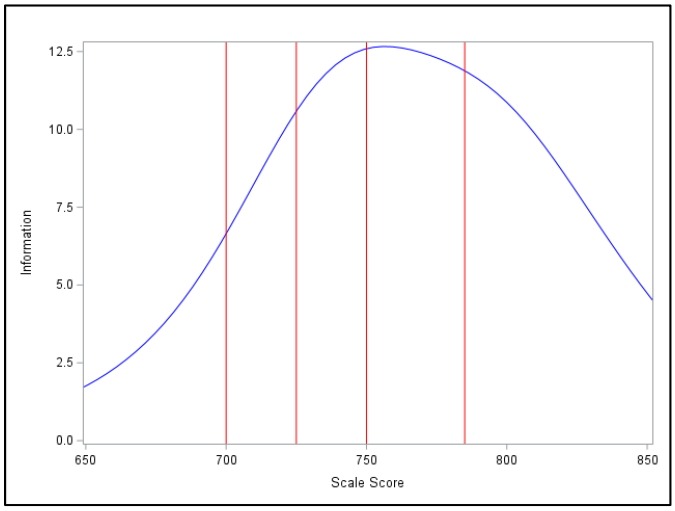


Figure N.33. ELA Grade 7 CSEM Curve

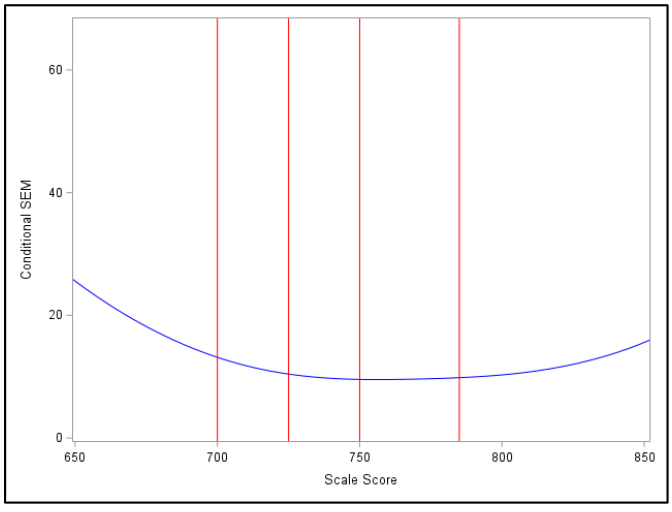


Figure N.34. ELA Grade 8 TCC

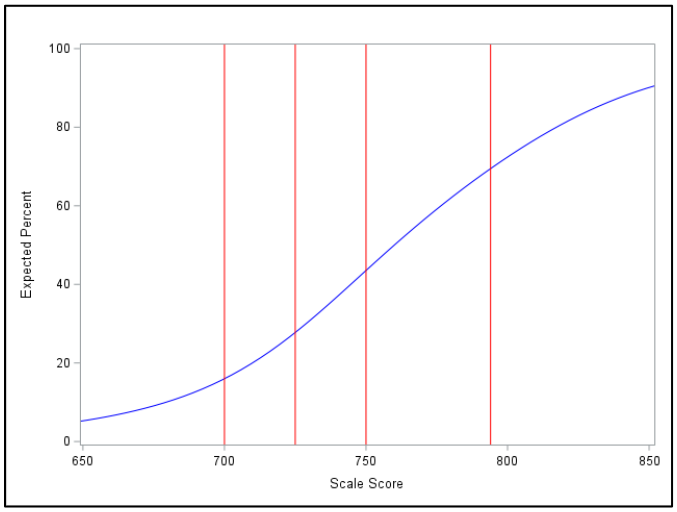


Figure N.35. ELA Grade 8 TIC

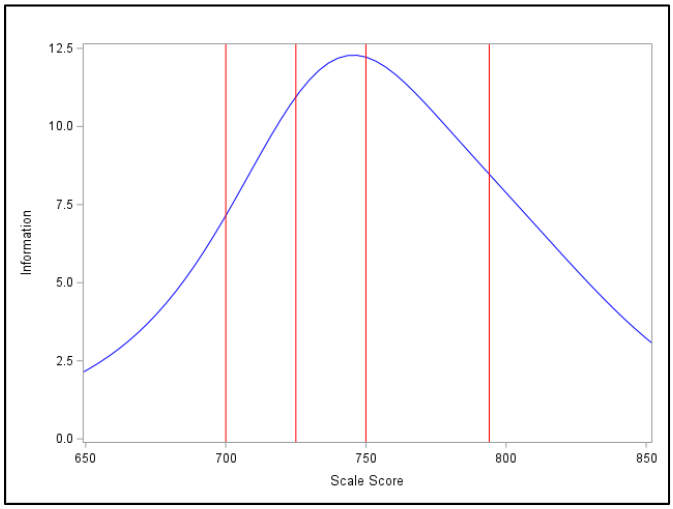


Figure N.36. ELA Grade 8 CSEM Curve

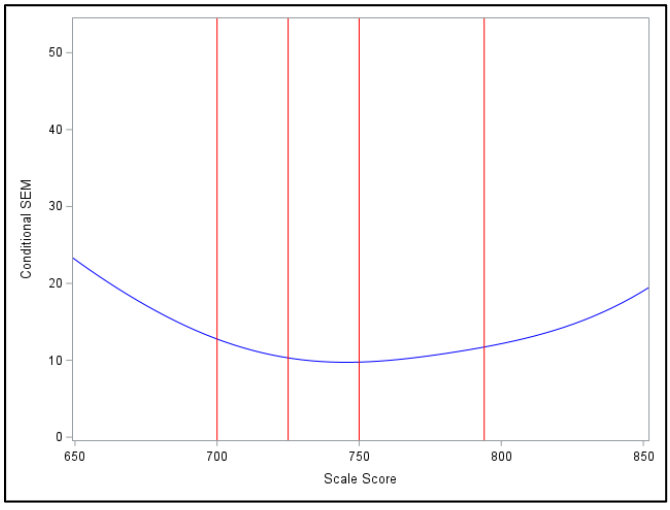


Figure N.37. CSLA Grade 3 TCC

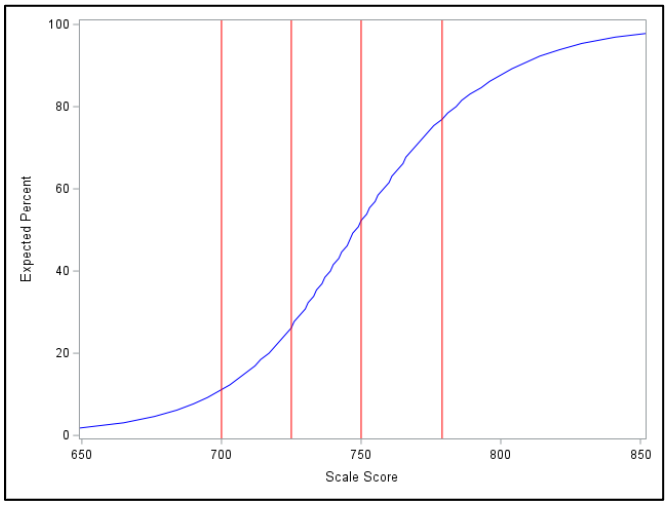


Figure N.38. CSLA Grade 3 TIC

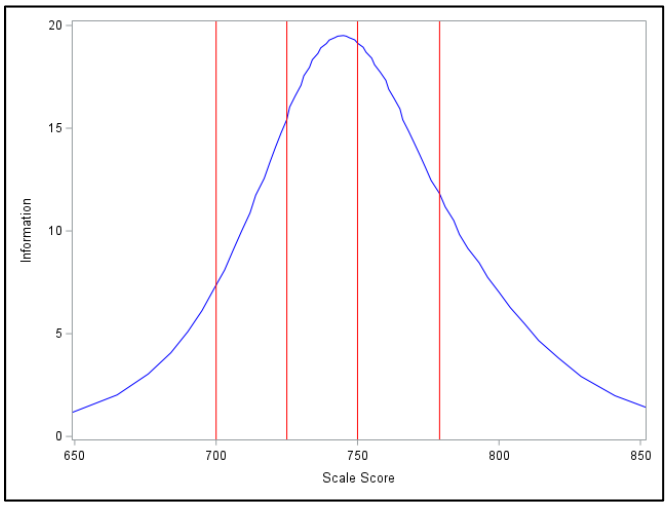


Figure N.39. CSLA Grade 3 CSEM Curve

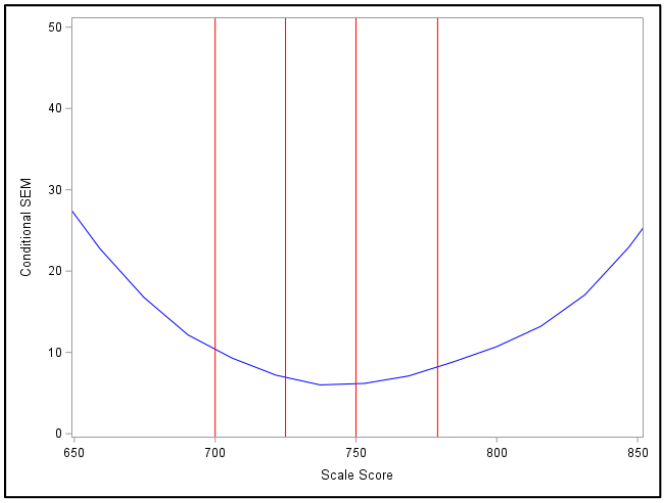


Figure N.40. CSLA Grade 4 TCC

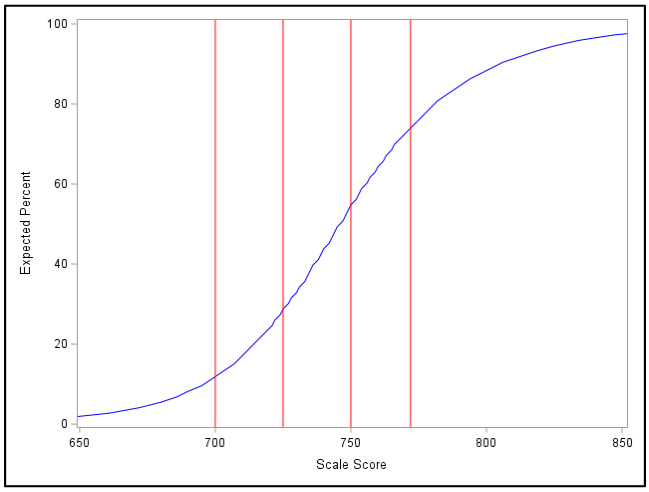


Figure N.41. CSLA Grade 4 TIC

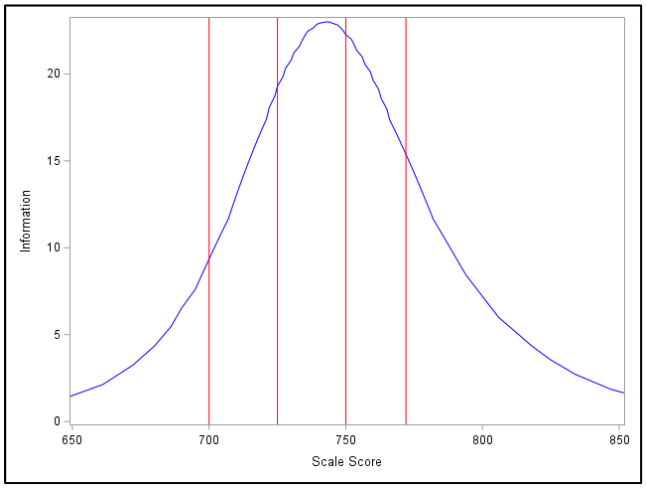


Figure N.42. CSLA Grade 4 CSEM Curve

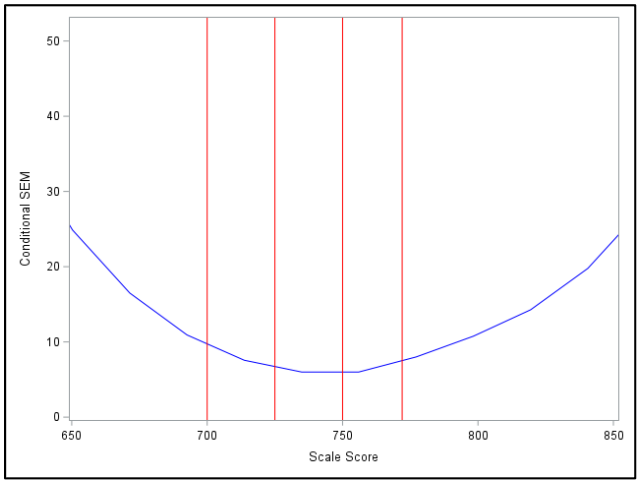


Figure N.43. ELA Reading Grade 3 TCC

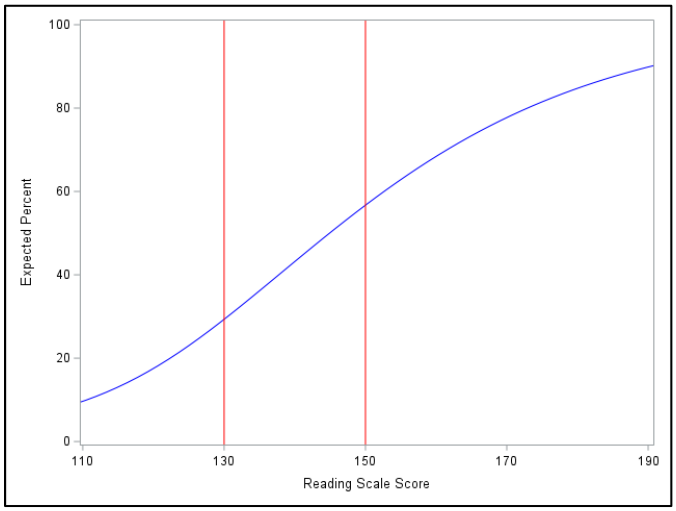


Figure N.44. ELA Reading Grade 3 TIC

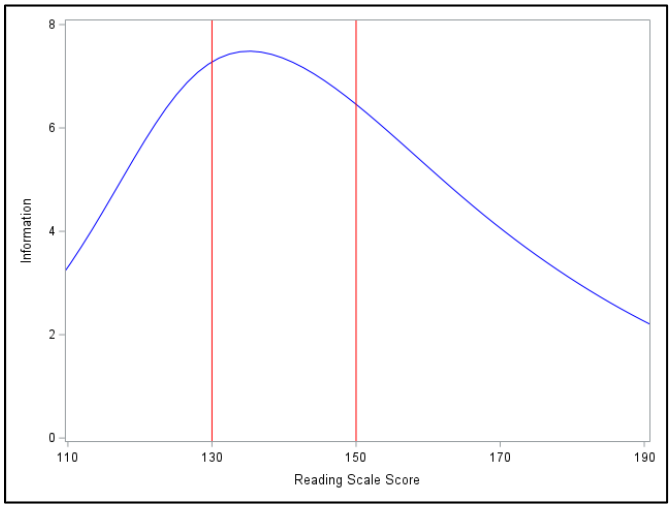


Figure N.45. ELA Reading Grade 3 CSEM Curve

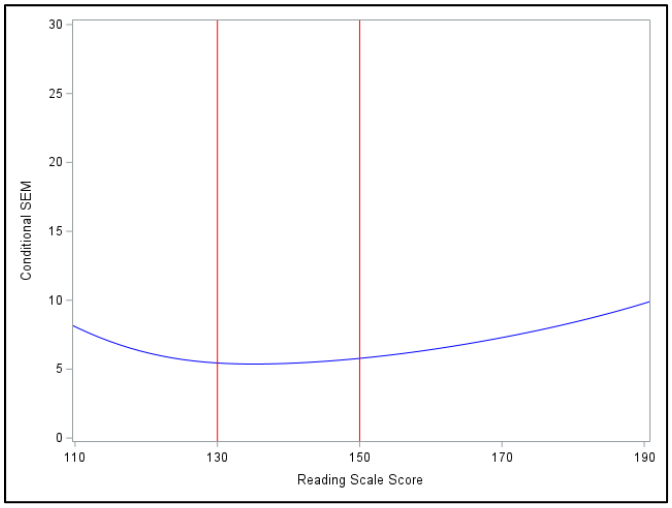


Figure N.46. ELA Reading Grade 4 TCC

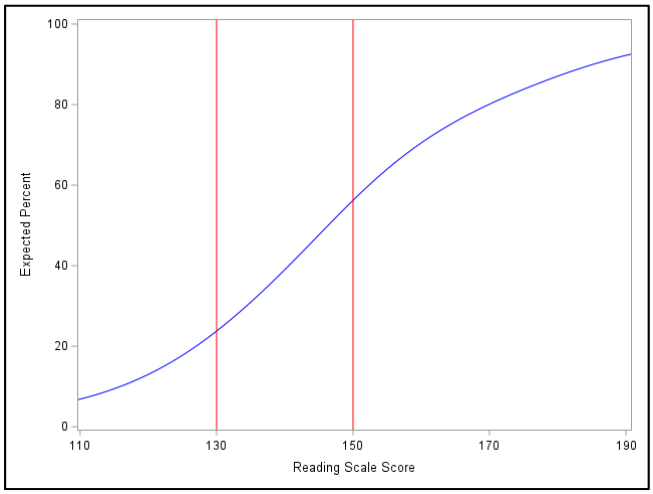


Figure N.47. ELA Reading Grade 4 TIC

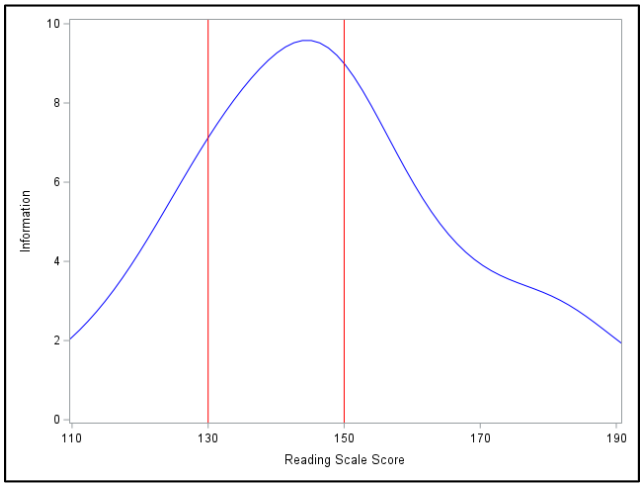


Figure N.48. ELA Reading Grade 4 CSEM Curve

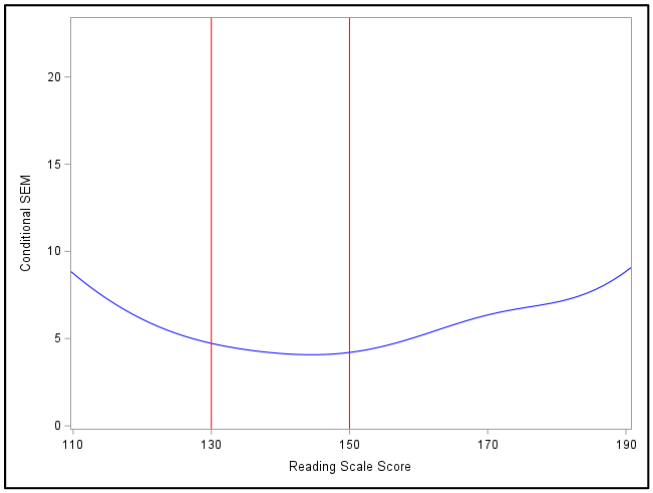


Figure N.49. ELA Reading Grade 5 TCC

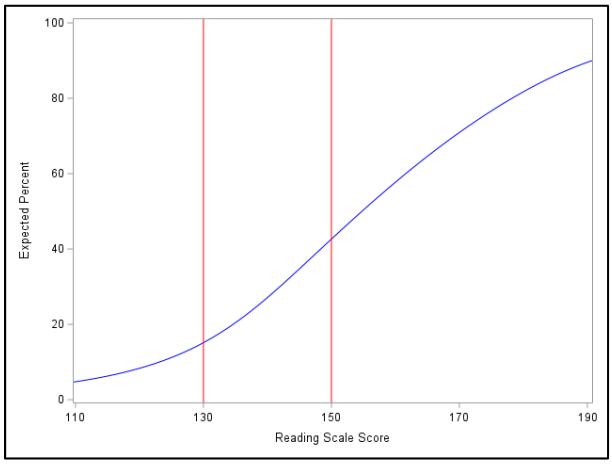


Figure N.50. ELA Reading Grade 5 TIC

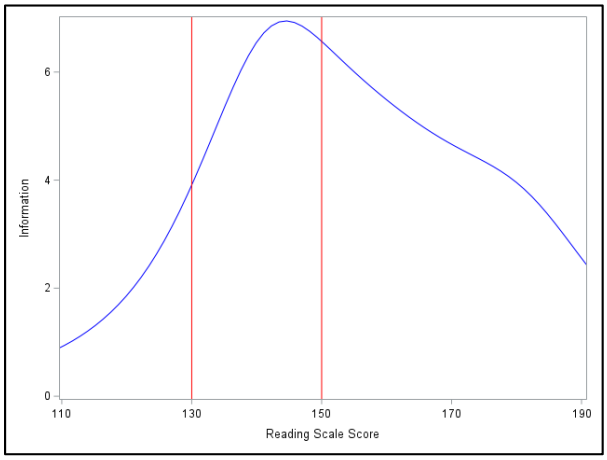


Figure N.51. ELA Reading Grade 5 CSEM Curve

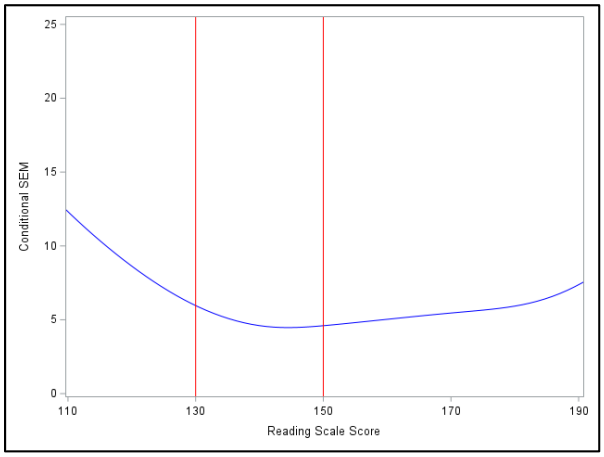


Figure N.52. ELA Reading Grade 6 TCC

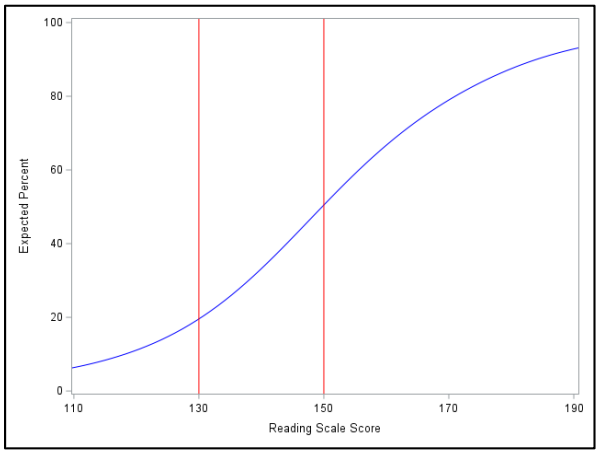


Figure N.53. ELA Reading Grade 6 TIC

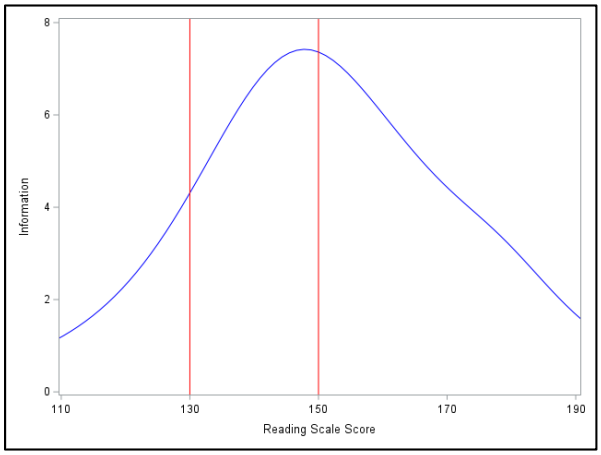


Figure N.54. ELA Reading Grade 6 CSEM Curve

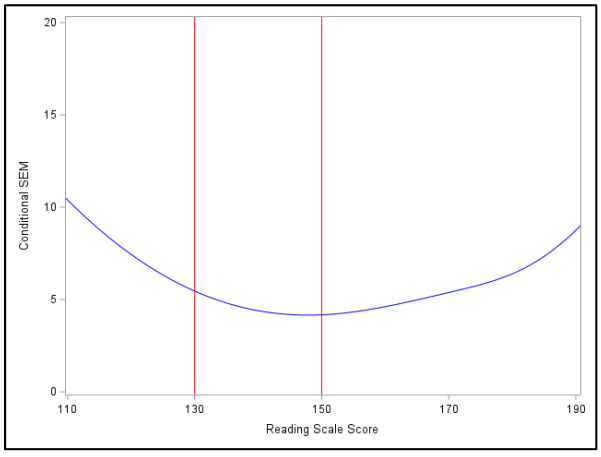


Figure N.55. ELA Reading Grade 7 TCC

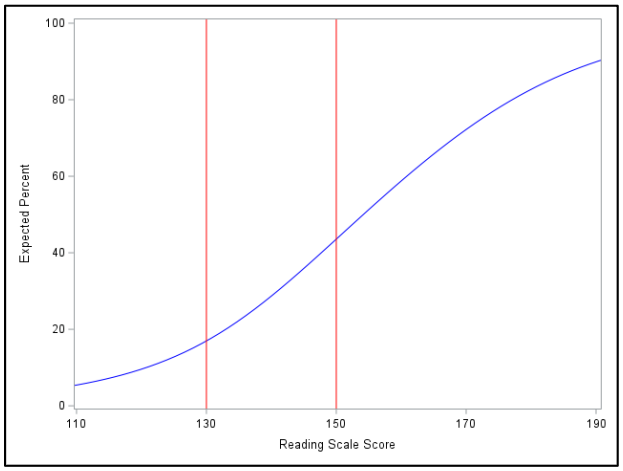


Figure N.56. ELA Reading Grade 7 TIC

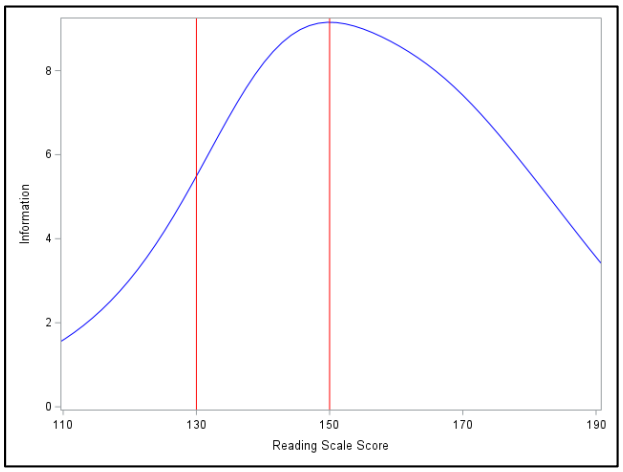


Figure N.57. ELA Reading Grade 7 CSEM Curve

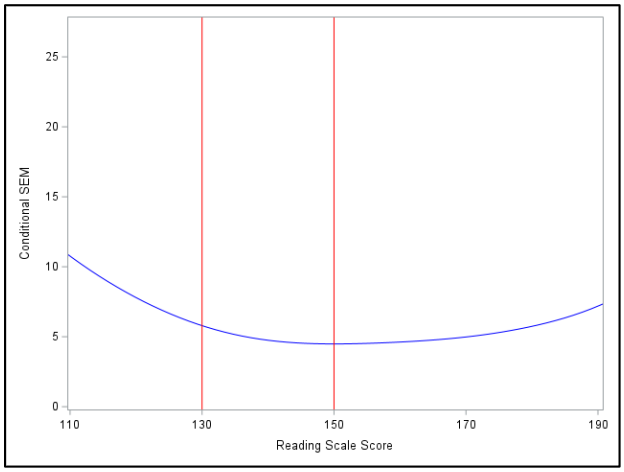


Figure N.58. ELA Reading Grade 8 TCC

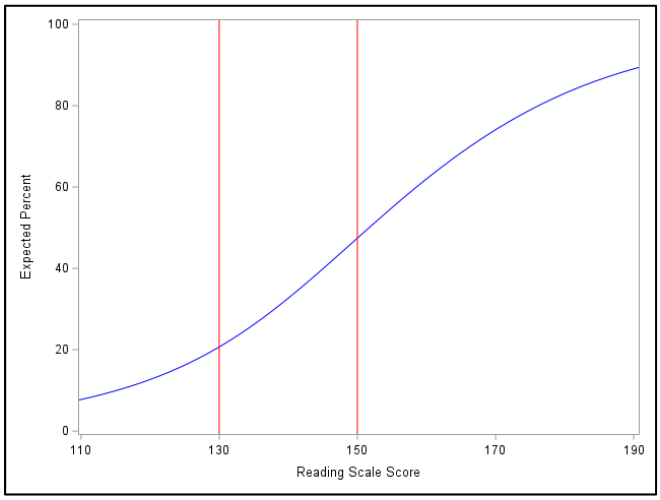


Figure N.59. ELA Reading Grade 8 TIC

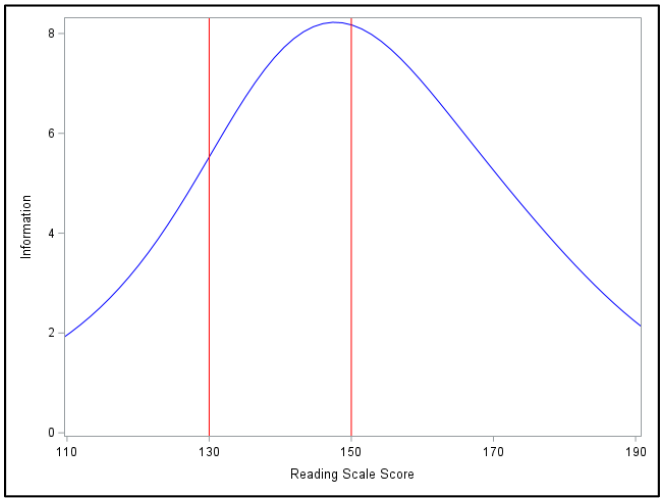


Figure N.60. ELA Reading Grade 8 CSEM Curve

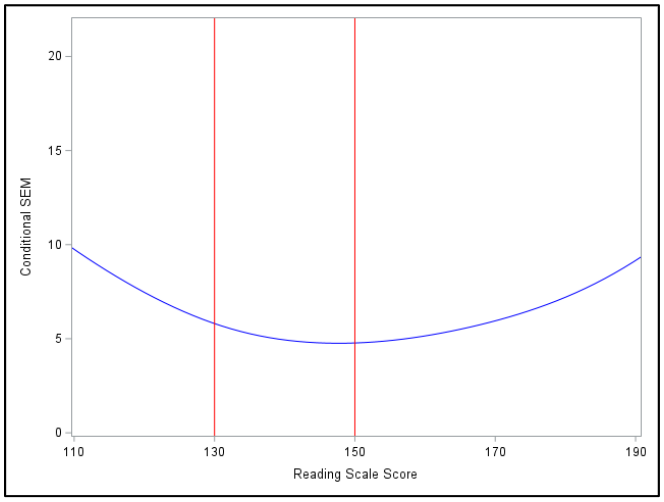


Figure N.61. CSLA Reading Grade 3 TCC

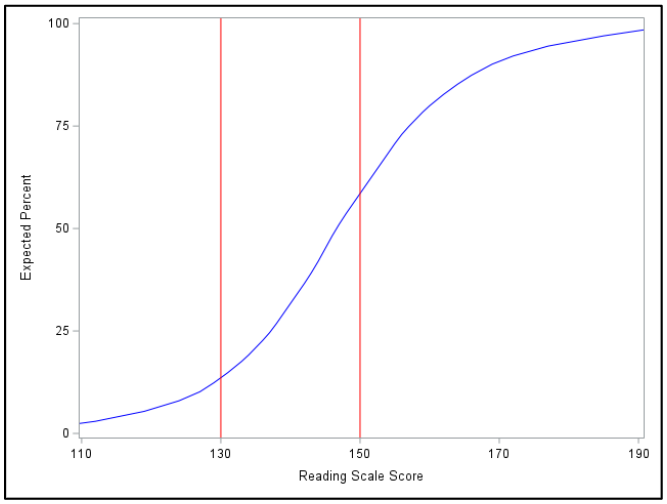


Figure N.62. CSLA Reading Grade 3 TIC

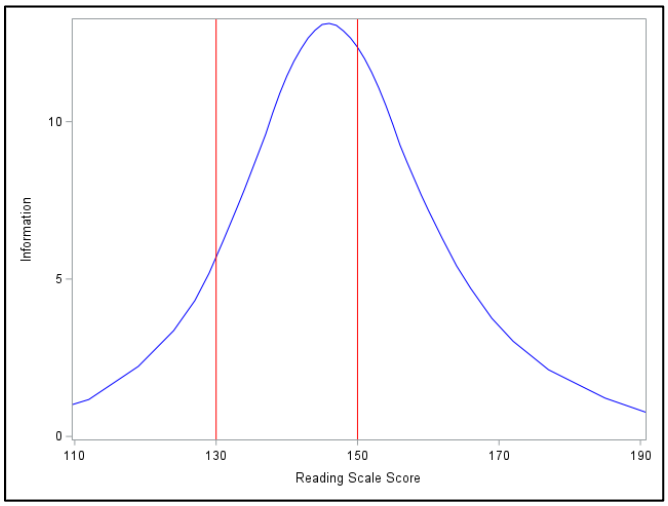


Figure N.63. CSLA Reading Grade 3 CSEM Curve

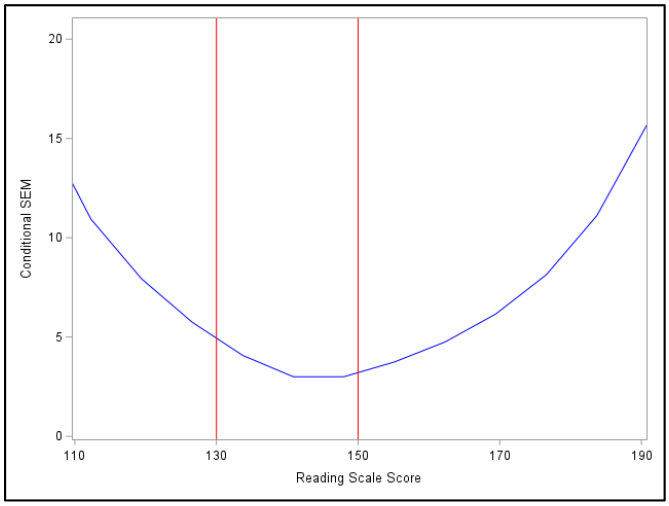


Figure N.64. CSLA Reading Grade 4 TCC

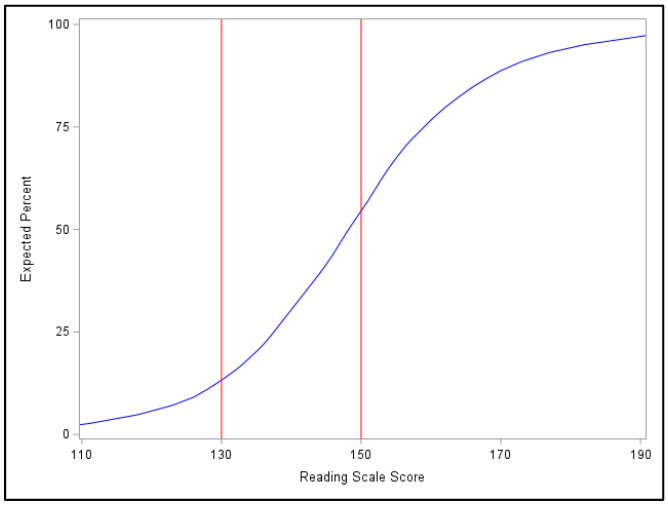


Figure N.65. CSLA Reading Grade 4 TIC

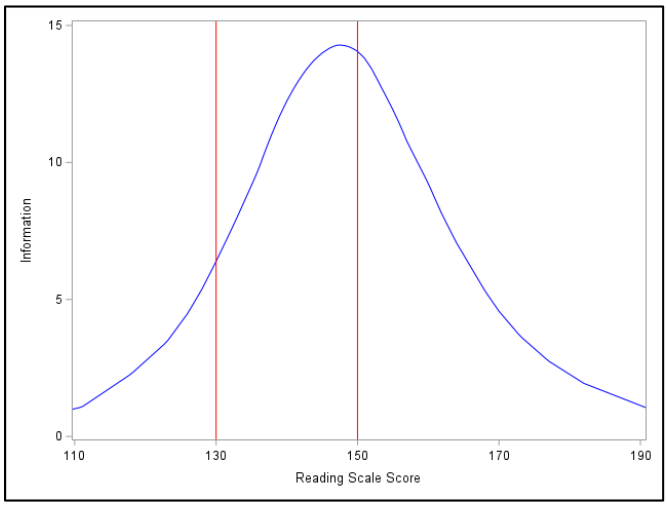
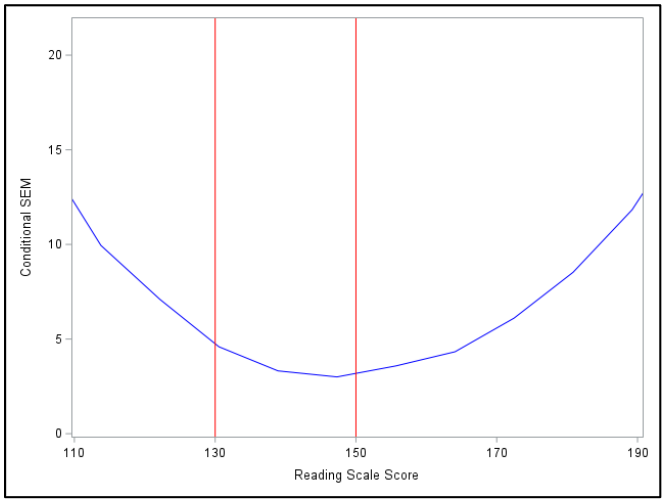


Figure N.66. CSLA Reading Grade 4 CSEM Curve



Appendix O: Inter-Rater Agreement

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	5,356	88.3	11.2	0.5	0.94	0	0.94
Item 2	4	5,441	83.5	15.9	0.6	0.91	0.01	0.91
Item 3, Part B	2	4,932	91.1	8.6	0.3	0.91	0	0.91
Item 4, Part B	3	4,937	83.8	15.1	1.1	0.92	0	0.92
Item 5, Part C	4	4,881	91.3	7.6	1.1	0.95	0	0.95

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	5,417	88.6	10.5	0.9	0.94	0	0.94
Item 2, Part B	3	4,986	93.2	5.8	1.0	0.97	0	0.97
Item 3	3	5,365	95.0	3.6	1.4	0.94	0	0.94
Item 4, Part	2	5,231	89.6	10.2	0.2	0.90	0	0.90
Item 4, Part B	2	5,059	93.5	6.4	0.1	0.94	0	0.94
Item 5	6	5,509	83.1	14.5	2.4	0.97	0	0.97

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, Part A	3	5,568	91.3	8.2	0.6	0.95	0	0.95
Item 1, Part B	3	5,323	93.5	5.1	1.4	0.95	0.01	0.95
Item 2	3	5,345	91.8	7.8	0.4	0.93	0	0.93
Item 3, Part	3	5,182	86.6	12.7	0.7	0.88	0.01	0.88
Item 4	4	5,493	80.8	17.5	1.7	0.93	0	0.93
Item 5, Part B	2	5,452	94.3	5.5	0.2	0.94	0	0.94

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, Part A	4	5,329	90.1	7.6	2.4	0.95	0	0.95
Item 1, Part B	2	5,194	96.3	3.6	0.1	0.94	0	0.94
Item 2	3	5,508	86.6	12.6	0.8	0.94	0	0.94
Item 3	4	5,392	84.5	13.4	2.1	0.90	0	0.90
Item 4	4	5,423	74.2	23.1	2.7	0.91	0.01	0.91
Item 5	3	5,470	81.8	16.6	1.6	0.89	0	0.89

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	5,262	88.2	11.3	0.6	0.94	0.00	0.94
Item 2	4	5,318	81.9	16.8	1.3	0.93	0.00	0.93
Item 3	3	5,270	83.6	15.5	0.8	0.82	0.01	0.82

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	5,014	92.2	7.4	0.4	0.96	0	0.96
Item 2, Part B	3	4,731	84.2	15.1	0.7	0.85	0.01	0.85
Item 3	4	4,873	88.7	10.3	0.9	0.89	0	0.89
Item 4	3	5,153	87.1	11.7	1.2	0.93	0	0.93
Item 5, Part A	4	4,967	93.5	5.7	0.8	0.96	0	0.96
Item 5, Part B	2	4,919	89.9	9.9	0.2	0.92	0.01	0.92

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, Part B	2	2,641	97.0	2.6	0.4	0.95	0.01	0.95
Item 2, Part B	2	2,616	95.6	4.4	0.1	0.96	0	0.96
Item 3, Part B	2	2,503	97.2	2.6	0.2	0.97	0	0.97
Item 4	3	3,355	95.6	4.3	0.1	0.98	0	0.98
Item 5, Part B	2	2,582	95.8	4.0	0.2	0.97	0	0.97
Item 6	3	2,574	75.8	21.9	2.3	0.85	0	0.85
Item 7	3	2,497	86.1	13.1	0.8	0.94	0	0.94
Item 8, Part B	2	2,597	92.4	7.2	0.4	0.93	0.01	0.93
Item 9, Part A	2	2,578	75.2	22.5	2.3	0.70	0.01	0.70
Item 9, Part B	2	2,451	87.6	12.3	0.1	0.87	0.01	0.87
Item 10, Part A	2	2,599	93.1	6.8	0.2	0.92	0.01	0.92
Item 10, Part B	2	2,543	89.5	10.4	0.1	0.87	0	0.87
Item 11	4	3,424	93.4	6.0	0.6	0.98	0	0.98
Item 12, Part B	3	2,637	91.5	7.8	0.7	0.96	0.01	0.96
Item 13, Part A	2	2,479	92.2	7.7	0.1	0.91	0	0.91
Item 13, Part B	2	2,356	85.7	14.1	0.1	0.86	0.01	0.86
Item 14, Part C	2	2,620	95.5	4.4	0.2	0.97	0	0.97
Item 15, Part A	2	2,670	89.9	10.1	0.1	0.90	0.01	0.90
Item 15, Part B	2	2,617	92.2	7.7	0.1	0.90	0	0.90
Item 16, Part B	3	2,576	87.4	11.1	1.5	0.91	0	0.91
Item 17, Part A	2	2,502	87.0	12.1	0.9	0.88	0.01	0.88
Item 17, Part B	2	2,452	95.1	4.9	0.0	0.95	0	0.95
Item 18, Part B	3	2,546	85.2	12.8	2.0	0.92	0	0.92
Item 19, Part B	2	2,617	94.2	5.4	0.4	0.96	0	0.96
Item 19, Part C	3	2,552	88.7	10.0	1.3	0.95	0	0.95
Item 20, Part B	2	2,568	95.1	4.2	0.8	0.94	0.01	0.94
Item 20, Part C	3	2,530	95.1	4.5	0.3	0.97	0.01	0.97
Item 21, Part B	3	2,661	93.8	5.9	0.3	0.96	0	0.96
Item 21, Part C	2	2,631	93.9	2.5	3.6	0.91	0	0.91
Item 22, Part A	2	2,591	90.2	9.4	0.4	0.92	0	0.92
Item 22, Part B	2	2,560	89.5	10.3	0.2	0.90	0.02	0.90
Item 23, Part B	2	2,438	91.7	8.2	0.2	0.88	0.01	0.88
Item 24, Part B	3	2,614	89.6	9.6	0.7	0.93	0.01	0.93
Item 25, Part B	3	2,613	92.9	6.4	0.7	0.92	0	0.92
Item 26, Part B	3	2,574	89.0	10.3	0.7	0.95	0.01	0.95

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, Part B	2	2,674	93.9	5.7	0.3	0.95	0	0.95
Item 2, Part B	2	2,719	94.3	5.6	0.2	0.95	0	0.95
Item 3, Part B	2	2,735	88.0	11.6	0.4	0.91	0	0.91
Item 4	3	2,733	91.7	7.3	1.1	0.89	0	0.89
Item 5, Part B	2	2,670	84.2	15.2	0.6	0.78	0	0.78
Item 6	3	2,687	91.0	8.5	0.5	0.96	0.01	0.96
Item 7	3	2,609	88.5	10.3	1.1	0.89	0	0.89
Item 8, Part A	2	2,623	89.7	10.3	0.1	0.88	0.02	0.88
Item 8, Part B	2	2,483	95.7	3.9	0.4	0.95	0	0.95
Item 9, Part A	2	2,685	84.4	15.1	0.6	0.86	0	0.86
Item 9, Part B	2	2,621	86.6	12.1	1.3	0.89	0	0.89
Item 10, Part A	2	2,688	95.6	4.4	0.0	0.97	0	0.97
Item 10, Part B	2	2,607	92.8	7.1	0.1	0.95	0.01	0.95
Item 11	4	2,706	88.6	10.7	0.7	0.96	0	0.96
Item 12, Part C	2	2,574	94.7	4.8	0.5	0.92	0	0.92
Item 13	4	2,681	83.0	15.5	1.6	0.92	0	0.92
Item 14, Part B	3	2,709	90.8	8.6	0.6	0.95	0	0.95
Item 15, Part A	2	2,701	83.8	16.1	0.1	0.86	0	0.86
Item 15, Part B	2	2,567	88.5	11.3	0.2	0.84	0.02	0.84
Item 16, Part A	2	2,688	88.6	11.3	0.1	0.89	0.01	0.89
Item 16, Part B	2	2,593	92.1	7.8	0.2	0.92	0	0.92
Item 17	3	2,803	73.0	25.0	2.0	0.72	0.02	0.72
Item 18	3	2,674	92.7	6.5	0.7	0.96	0.01	0.96
Item 19, Part B	3	2,602	84.2	14.8	0.9	0.89	0	0.89
Item 20, Part A	2	2,754	84.8	14.9	0.3	0.87	0.02	0.87
Item 20, Part B	2	2,677	80.8	18.9	0.3	0.82	0.03	0.82
Item 21, Part A	2	2,664	85.0	14.7	0.3	0.84	0	0.84
Item 21, Part B	2	2,579	86.5	13.1	0.3	0.86	0.01	0.86
Item 22	4	2,723	76.9	20.2	2.8	0.90	0.01	0.90
Item 23, Part A	2	2,693	88.4	11.3	0.3	0.88	0	0.88
Item 23, Part B	2	2,599	92.3	7.6	0.1	0.89	0.01	0.89
Item 24, Part A	2	2,884	93.6	6.1	0.3	0.95	0	0.95
Item 24, Part B	2	2,730	91.0	8.9	0.1	0.89	0.01	0.89
Item 25, Part B	3	2,669	88.0	9.6	2.4	0.81	0.03	0.81
Item 26, Part B	3	2,688	77.6	21.3	1.1	0.89	0.01	0.89
Item 27	3	3,923	95.8	4.1	0.1	0.97	0.01	0.97
Item 28	3	2,693	94.9	3.4	1.7	0.97	0	0.97

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	4	2,745	90.3	8.5	1.2	0.96	0	0.96
Item 2, Part B	2	2,949	96.1	3.9	0	0.95	0.01	0.95
Item 3, Part B	2	2,659	96.8	2.0	1.2	0.92	0.01	0.92
Item 4, Part B	2	2,862	95.8	4.0	0.1	0.96	0	0.96
Item 5, Part B	2	2,843	95.5	4.1	0.4	0.87	0.01	0.87
Item 6	3	2,818	91.6	6.5	1.8	0.94	0	0.94
Item 7	3	2,861	89.3	10.3	0.4	0.94	0.01	0.94
Item 8, Part A	2	2,834	93.2	6.6	0.2	0.92	0	0.92
Item 9	3	2,752	83.6	15.0	1.5	0.92	0.01	0.92
Item 10	3	2,721	87.2	12.5	0.3	0.92	0	0.92
Item 11	3	2,912	95.0	4.8	0.1	0.97	0	0.97
Item 12	4	2,771	88.6	9.5	1.9	0.94	0	0.94
Item 13	4	2,796	88.5	10.2	1.3	0.96	0	0.96
Item 14, Part A	2	2,851	90.9	9.0	0.1	0.91	0	0.91
Item 14, Part B	2	2,806	92.1	7.9	0.0	0.92	0.01	0.92
Item 15, Part A	2	2,863	90.5	9.4	0.1	0.91	0.01	0.91
Item 15, Part B	2	2,824	90.1	9.8	0.1	0.91	0.01	0.91
Item 16, Part A	2	2,824	91.8	8.1	0.1	0.93	0	0.93
Item 16, Part B	2	2,773	88.5	11.1	0.4	0.91	0.01	0.91
Item 17, Part A	2	2,789	77.4	21.6	1.0	0.77	0.01	0.77
Item 17, Part B	2	2,746	91.2	8.6	0.2	0.91	0.01	0.91
Item 18, Part A	2	2,816	96.6	3.4	0.0	0.96	0.01	0.96
Item 18, Part B	2	2,711	97.6	2.3	0.1	0.96	0	0.96
Item 19, Part A	2	2,984	96.8	3.2	0.0	0.97	0	0.97
Item 19, Part B	2	2,946	96.3	3.7	0.0	0.95	0	0.95
Item 20	4	2,909	84.3	14.6	1.1	0.95	0	0.95
Item 21, Part A	2	2,795	85.7	14.0	0.4	0.87	0.01	0.87
Item 21, Part B	2	2,726	81.4	18.2	0.4	0.82	0.01	0.82
Item 22, Part A	2	2,883	92.2	7.6	0.2	0.95	0	0.95
Item 22, Part B	2	2,834	92.2	7.5	0.2	0.94	0.01	0.94
Item 23, Part A	3	2,737	80.9	16.9	2.2	0.86	0.04	0.86
Item 23, Part B	3	2,685	89.8	9.2	0.9	0.96	0.01	0.96
Item 24, Part A	2	2,850	87.6	12.3	0.1	0.89	0.01	0.89
Item 24, Part B	2	2,821	93.4	6.3	0.3	0.92	0	0.92
Item 24, Part C	2	2,796	84.0	15.9	0.1	0.89	0.01	0.89
Item 25	4	2,768	85.7	11.7	2.6	0.95	0.01	0.95
Item 26	4	2,920	80.0	18.4	1.6	0.93	0	0.93
Item 27	4	2,779	85.5	13.7	0.8	0.96	0	0.96

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, Part B	2	3,052	89.8	10.0	0.2	0.87	0	0.87
Item 2	3	2,975	84.0	14.1	1.9	0.82	0.01	0.82
Item 3, Part B	2	2,661	92.8	7.0	0.2	0.95	0	0.95
Item 4	3	2,980	82.3	17.3	0.3	0.86	0.01	0.86
Item 5, Part B	3	2,926	94.3	5.3	0.5	0.94	0	0.94
Item 6	4	2,962	86.1	13.8	0.1	0.91	0.01	0.91
Item 7, Part C	2	2,953	93.8	6.2	0.0	0.92	0	0.92
Item 8	4	2,984	88.3	10.9	0.8	0.95	0	0.95
Item 9	3	2,996	80.9	17.0	2.1	0.81	0	0.81
Item 10, Part B	3	2,338	90.2	9.6	0.3	0.91	0.01	0.91
Item 11, Part A	2	2,902	95.5	4.5	0.0	0.97	0	0.97
Item 11, Part B	2	2,881	95.3	4.6	0.1	0.97	0	0.97
Item 12, Part A	2	2,931	92.1	7.9	0.0	0.91	0.01	0.91
Item 12, Part B	2	2,852	97.7	2.3	0.0	0.94	0	0.94
Item 13	4	2,937	85.6	14.3	0.1	0.95	0	0.95
Item 14, Part A	3	2,987	84.1	15.7	0.1	0.95	0	0.95
Item 14, Part B	3	2,882	86.6	13.4	0.0	0.93	0.01	0.93
Item 15	4	2,946	85.3	13.9	0.8	0.90	0.01	0.90
Item 16, Part A	2	2,992	91.6	8.0	0.4	0.95	0.01	0.95
Item 16, Part B	2	2,953	93.7	6.0	0.3	0.95	0	0.95
Item 17	4	2,956	87.8	12	0.1	0.95	0	0.95
Item 18, Part A	2	2,943	91.5	8.5	0.0	0.94	0	0.94
Item 18, Part B	2	2,867	90.6	9.4	0.0	0.84	0.01	0.84
Item 19, Part A	3	2,896	88.5	10.5	1.1	0.91	0.01	0.91
Item 19, Part B	3	2,795	93.1	6.8	0.1	0.89	0.01	0.89
Item 20, Part B	2	3,001	96.3	3.4	0.3	0.97	0.01	0.97
Item 20, Part C	3	2,917	86.1	11.8	2.1	0.92	0	0.92

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	2,910	86.2	12.6	1.1	0.92	0	0.92
Item 2	3	2,884	92.3	7.1	0.6	0.95	0.01	0.95
Item 3	3	2,862	95.5	4.3	0.2	0.98	0.01	0.98
Item 4	4	2,891	85.8	11.8	2.4	0.92	0.01	0.92
Item 5	3	2,954	91.3	8.1	0.6	0.95	0.01	0.95
Item 6	4	2,904	89.2	10.5	0.3	0.95	0.01	0.95
Item 7	3	2,972	92.4	6.3	1.3	0.93	0	0.93
Item 8	4	2,947	82.4	15.2	2.3	0.95	0	0.95

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1, Part A	2	2,908	91.5	8.1	0.4	0.86	0.01	0.86
Item 2, Part A	2	2,864	92.5	7.2	0.2	0.90	0.02	0.90
Item 2, Part B	2	2,799	90.9	8.7	0.4	0.82	0.02	0.82
Item 3, Part B	2	2,892	91.8	6.2	2.0	0.88	0	0.88
Item 4	3	2,884	94.9	4.8	0.3	0.97	0	0.97
Item 5, Part A	2	2,849	95.5	4.5	0.1	0.95	0	0.95
Item 6	3	2,926	93.8	5.3	1.0	0.83	0	0.83
Item 7	4	2,905	79.7	18.0	2.3	0.90	0	0.90
Item 8, Part A	2	2,943	87.8	11.9	0.3	0.84	0.01	0.84
Item 8, Part B	2	2,897	89.7	10.2	0.1	0.86	0	0.86
Item 9, Part A	3	3,005	82.5	15.3	2.2	0.86	0.01	0.86
Item 9, Part B	3	2,935	95.6	4.0	0.4	0.94	0.01	0.94
Item 10, Part B	2	2,920	86.5	10.9	2.7	0.73	0.01	0.73
Item 11, Part C	2	2,876	87.4	12.2	0.3	0.83	0	0.83
Item 12, Part B	3	2,885	85.2	13.8	0.9	0.86	0.01	0.86
Item 13, Part A	2	2,887	95.8	3.8	0.3	0.95	0	0.95
Item 13, Part B	2	2,830	93.6	5.9	0.5	0.90	0.01	0.90
Item 14, Part A	2	2,845	95.2	3.8	1.0	0.93	0.01	0.94
Item 14, Part B	2	2,792	99.0	0.9	0.1	0.91	0	0.91
Item 15, Part A	3	3,004	94.9	4.2	0.9	0.96	0	0.96
Item 15, Part B	3	2,905	86.5	10.6	2.9	0.88	0	0.88
Item 16	3	3,029	82.6	16.6	0.9	0.88	0.01	0.88
Item 17	3	2,935	81.6	17.6	0.8	0.93	0.01	0.93
Item 18, Part A	2	2,893	80.1	19.5	0.4	0.73	0	0.73
Item 18, Part B	2	2,804	82.5	17.0	0.6	0.73	0	0.73
Item 19, Part A	2	2,869	95.5	4.2	0.3	0.97	0	0.97
Item 19, Part B	2	2,773	97.2	2.6	0.2	0.96	0	0.96
Item 20, Part B	3	2,957	84.8	14.5	0.7	0.91	0.01	0.91
Item 21	4	3,031	92.5	6.4	1.1	0.96	0.01	0.96
Item 22	3	2,949	89.0	10.6	0.3	0.93	0.01	0.93
Item 23, Part A	2	2,852	89.8	10.0	0.2	0.88	0.01	0.88
Item 23, Part B	2	2,816	91.5	8.3	0.1	0.92	0.01	0.92

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	3	5,522	82.0	18.0	0.0	0.82	0.00	0.82
PCR 1 WKL	3	5,522	86.4	13.6	0.0	0.83	0.01	0.83
PCR 2 WE	3	5,524	84.1	15.8	0.1	0.80	0.02	0.80
PCR 2 WKL	3	5,524	84.1	15.8	0.1	0.80	0.02	0.80

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	3	5,571	80.6	19.4	0.0	0.87	0.02	0.87
PCR 1 WKL	3	5,571	81.2	18.8	0.0	0.87	0.06	0.87
PCR 2 WE	4	5,577	76.3	23.4	0.3	0.73	0.01	0.73
PCR 2 WKL	3	5,577	76.3	23.4	0.3	0.73	0.01	0.73

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	3	5,744	83.3	16.4	0.3	0.87	0.01	0.87
PCR 1 WKL	3	5,744	82.8	16.8	0.4	0.87	0.00	0.87
PCR 2 WE	4	5,745	75.9	23.7	0.4	0.73	0.00	0.73
PCR 2 WKL	3	5,745	75.9	23.7	0.4	0.73	0.00	0.73

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	4	5,607	81.6	18.0	0.5	0.86	0.01	0.86
PCR 1 WKL	3	5,607	81.6	18.0	0.5	0.86	0.01	0.86
PCR 2 WE	4	5,600	83.0	17.0	0.0	0.89	0.01	0.89
PCR 2 WKL	3	5,600	80.2	19.8	0.0	0.88	0.02	0.88

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	4	5,531	81.2	18.8	0.0	0.86	0.02	0.86
PCR 1 WKL	3	5,531	81.2	18.8	0.0	0.86	0.02	0.86
PCR 2 WE	4	5,539	79.6	19.4	0.9	0.89	0.01	0.89
PCR 2 WKL	3	5,539	77.8	21.3	1.0	0.89	0.00	0.89

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	4	5,287	79.2	20.7	0.1	0.90	0.02	0.90
PCR 1 WKL	3	5,287	78.0	22.0	0.0	0.91	0.03	0.91
PCR 2 WE	4	5,303	76.2	23.8	0.0	0.90	0.00	0.90
PCR 2 WKL	3	5,303	76.2	23.8	0.0	0.90	0.00	0.90

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	3	4,172	82.7	16.9	0.3	0.80	0	0.80
PCR 1 WKL	3	4,172	79.7	20.1	0.1	0.74	0.03	0.74
PCR 2 WE	3	8,295	83.4	16.6	0.0	0.74	0.01	0.74
PCR 2 WKL	3	8,295	87.6	12.4	0.0	0.70	0.01	0.70
PCR 3 WE	3	8,468	87.2	12.8	0.0	0.80	0	0.80
PCR 3 WKL	3	8,468	84.1	15.9	0.0	0.64	0.01	0.64
PCR 4 WE	3	4,244	86.9	13.1	0.0	0.78	0	0.78
PCR 4 WKL	3	4,244	83.7	16.3	0.0	0.63	0.02	0.63
PCR 5 WE	3	4,118	86.4	13.5	0.1	0.82	0	0.82
PCR 5 WKL	3	4,118	85.9	13.6	0.4	0.83	0	0.83
PCR 6 WE	3	4,224	86.9	13.0	0.0	0.83	0	0.83
PCR 6 WKL	3	4,224	82.7	16.7	0.7	0.75	0.02	0.75
PCR 7 WE	3	4,215	85.6	14.4	0.0	0.78	0	0.78
PCR 7 WKL	3	4,215	85.6	14.4	0.0	0.78	0	0.78
PCR 8 WE	3	4,167	82.6	17.3	0.1	0.79	0.02	0.79
PCR 8 WKL	3	4,167	82.6	17.3	0.1	0.79	0.02	0.79
PCR 9 WE	3	4,192	81.4	18.6	0.0	0.74	0.01	0.74
PCR 9 WKL	3	4,192	85.3	14.7	0.0	0.74	0.02	0.75
PCR 10 WE	3	4,368	81.9	18.1	0.0	0.72	0	0.72
PCR 10 WKL	3	4,368	84.8	15.2	0.0	0.71	0.01	0.71

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	3	4,344	80.4	19.0	0.7	0.80	0.02	0.80
PCR 1 WKL	3	4,344	83.5	16.3	0.3	0.87	0.01	0.87
PCR 2 WE	3	4,319	79.8	19.8	0.4	0.78	0.01	0.78
PCR 2 WKL	3	4,319	83.8	15.9	0.3	0.84	0.01	0.84
PCR 3 WE	3	4,220	79.7	20.3	0.0	0.84	0	0.84
PCR 3 WKL	3	4,220	81.5	18.5	0.0	0.87	0.01	0.87
PCR 4 WE	3	4,311	80.2	19.8	0.0	0.84	0	0.84
PCR 4 WKL	3	4,311	83.2	16.8	0.0	0.88	0.01	0.88
PCR 5 WE	3	8,482	80.5	19.1	0.5	0.80	0.01	0.80
PCR 5 WKL	3	8,482	80.0	19.6	0.4	0.82	0.02	0.82
PCR 6 WE	3	4,291	81.3	18.7	0.0	0.82	0.01	0.82
PCR 6 WKL	3	4,291	80.6	19.4	0.0	0.83	0.01	0.83
PCR 7 WE	3	4,291	81.3	18.2	0.5	0.79	0.01	0.79
PCR 7 WKL	3	4,291	81.3	18.3	0.4	0.82	0	0.82
PCR 8 WE	3	4,319	79.7	19.8	0.5	0.71	0	0.71
PCR 8 WKL	3	4,319	88.5	10.7	0.8	0.82	0	0.82
PCR 9 WE	4	8,486	73.5	26.1	0.4	0.71	0.02	0.71
PCR 9 WKL	3	8,486	73.5	26.1	0.4	0.71	0.02	0.71
PCR 10 WE	4	4,391	74.9	24.7	0.3	0.72	0.01	0.72
PCR 10 WKL	3	4,391	74.9	24.7	0.3	0.72	0.01	0.72

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	3	4,408	80.5	19.4	0.1	0.87	0.01	0.87
PCR 1 WKL	3	4,408	82.4	16.8	0.8	0.83	0.01	0.83
PCR 2 WE	3	4,380	80.6	19.4	0.0	0.89	0.01	0.89
PCR 2 WKL	3	4,380	79.8	20.1	0.0	0.89	0.01	0.89
PCR 3 WE	3	4,454	82.1	17.7	0.2	0.87	0.01	0.87
PCR 3 WKL	3	4,454	81.7	18.2	0.2	0.88	0.01	0.88
PCR 4 WE	3	8,963	81.3	18.6	0.1	0.90	0.01	0.90
PCR 4 WKL	3	8,963	81.2	18.8	0.0	0.90	0	0.90
PCR 5 WE	3	4,416	82.2	17.7	0.1	0.85	0	0.85
PCR 5 WKL	3	4,416	82.2	17.3	0.5	0.86	0	0.86
PCR 6 WE	3	4,499	80.9	19.1	0.0	0.85	0	0.85
PCR 6 WKL	3	4,499	81.4	18.5	0.0	0.86	0.01	0.86
PCR 7 WE	3	8,845	81.5	18.3	0.2	0.89	0	0.89
PCR 7 WKL	3	8,845	80.3	19.5	0.2	0.88	0	0.88
PCR 8 WE	3	4,431	79.7	20.3	0.0	0.83	0	0.83
PCR 8 WKL	3	4,431	81.8	18.2	0.0	0.87	0.02	0.87
PCR 9 WE	4	4,439	73.4	25.9	0.7	0.79	0	0.79
PCR 9 WKL	3	4,439	73.4	25.9	0.7	0.79	0	0.79
PCR 10 WE	4	4,430	78.2	21.4	0.4	0.82	0.01	0.82
PCR 10 WKL	3	4,430	78.2	21.4	0.4	0.82	0.01	0.82

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	4	3,688	81.3	18.7	0.0	0.88	0.03	0.88
PCR 1 WKL	3	3,688	75.6	24.4	0.0	0.88	0.01	0.88
PCR 2 WE	4	3,697	81.3	18.7	0.0	0.89	0.07	0.89
PCR 2 WKL	3	3,697	74.3	25.7	0.0	0.88	0.05	0.88
PCR 3 WE	4	3,766	80.6	19.4	0.0	0.85	0	0.85
PCR 3 WKL	3	3,766	80.6	19.4	0.0	0.85	0.01	0.85
PCR 4 WE	4	3,755	80.4	19.6	0.0	0.88	0.02	0.88
PCR 4 WKL	3	3,755	77.7	22.2	0.0	0.91	0.01	0.91
PCR 5 WE	4	7,664	81.7	18.0	0.3	0.87	0.01	0.87
PCR 5 WKL	3	7,664	81.7	18.0	0.3	0.87	0.01	0.87
PCR 6 WE	4	3,704	75.6	23.6	0.7	0.83	0.02	0.83
PCR 6 WKL	3	3,704	75.6	23.6	0.7	0.83	0.02	0.83
PCR 7 WE	4	3,681	75.7	23.8	0.5	0.83	0.01	0.83
PCR 7 WKL	3	3,681	75.7	23.8	0.5	0.83	0.01	0.83
PCR 8 WE	4	3,725	80.1	19.7	0.2	0.85	0.02	0.85
PCR 8 WKL	3	3,725	80.1	19.7	0.2	0.85	0.02	0.85
PCR 9 WE	4	3,728	80.9	19.0	0.1	0.86	0.01	0.86
PCR 9 WKL	3	3,728	80.5	19.3	0.1	0.89	0.01	0.89
PCR 10 WE	4	7,573	79.8	20.1	0.0	0.90	0	0.90
PCR 10 WKL	3	7,573	81.1	18.8	0.1	0.92	0	0.92
PCR 11 WE	4	3,752	78.6	21.2	0.1	0.83	0	0.83
PCR 11 WKL	3	3,752	78.6	21.2	0.1	0.83	0	0.83
PCR 12 WE	4	3,841	82.3	17.6	0.1	0.87	0	0.87
PCR 12 WKL	3	3,841	82.3	17.6	0.1	0.87	0	0.87

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	4	7,457	76.2	23.8	0.0	0.83	0.02	0.83
PCR 1 WKL	3	7,457	76.2	23.8	0.0	0.83	0.02	0.83
PCR 2 WE	4	3,769	72.4	27.5	0.0	0.79	0.01	0.79
PCR 2 WKL	3	3,769	72.4	27.5	0.0	0.79	0.01	0.79
PCR 3 WE	4	7,358	77.7	21.9	0.4	0.86	0	0.86
PCR 3 WKL	3	7,358	77.7	21.9	0.4	0.86	0	0.86
PCR 4 WE	4	3,700	78.3	21.6	0.1	0.86	0.02	0.86
PCR 4 WKL	3	3,700	78.3	21.6	0.1	0.86	0.02	0.86
PCR 5 WE	4	3,724	77.8	22.2	0.1	0.83	0.01	0.83
PCR 5 WKL	3	3,724	77.8	22.2	0.1	0.83	0.01	0.83
PCR 6 WE	4	3,614	75.0	25.0	0.0	0.83	0.01	0.83
PCR 6 WKL	3	3,614	75.0	25.0	0.0	0.83	0.01	0.83
PCR 7 WE	4	3,759	74.8	25.1	0.0	0.84	0.03	0.84
PCR 7 WKL	3	3,759	74.8	25.1	0.0	0.84	0.03	0.84
PCR 8 WE	4	3,697	74.1	25.9	0.0	0.82	0.01	0.82
PCR 8 WKL	3	3,697	74.1	25.9	0.0	0.82	0.01	0.82
PCR 9 WE	4	3,728	80.3	19.6	0.1	0.87	0	0.87
PCR 9 WKL	3	3,728	79.5	20.4	0.1	0.87	0	0.87
PCR 10 WE	4	3,739	80.2	19.2	0.6	0.85	0	0.85
PCR 10 WKL	3	3,739	83.3	16.3	0.5	0.89	0	0.89
PCR 11 WE	4	3,820	77.6	22.4	0.0	0.86	0.01	0.86
PCR 11 WKL	3	3,820	77.6	22.4	0.0	0.86	0.01	0.86
PCR 12 WE	4	3,688	75.6	24.4	0.0	0.86	0.01	0.86
PCR 12 WKL	3	3,688	75.6	24.4	0.0	0.86	0.01	0.86

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	4	3,557	75.5	24.2	0.4	0.89	0	0.89
PCR 1 WKL	3	3,557	75.5	24.2	0.4	0.89	0	0.89
PCR 2 WE	4	3,439	74.7	25.3	0.0	0.84	0.02	0.84
PCR 2 WKL	3	3,439	74.7	25.3	0.0	0.84	0.02	0.84
PCR 3 WE	4	3,666	79.7	20.0	0.3	0.88	0	0.88
PCR 3 WKL	3	3,666	80.8	19.1	0.1	0.89	0.01	0.89
PCR 4 WE	4	3,670	79.5	19.8	0.7	0.87	0.01	0.87
PCR 4 WKL	3	3,670	74.8	24.5	0.7	0.86	0.01	0.86
PCR 5 WE	4	3,597	79.6	19.1	1.3	0.88	0	0.88
PCR 5 WKL	3	3,597	77.6	21.2	1.1	0.89	0.01	0.89
PCR 6 WE	4	7,094	79.5	20.2	0.3	0.88	0	0.88
PCR 6 WKL	3	7,094	78.8	21.1	0.1	0.87	0.01	0.87
PCR 7 WE	4	3,501	79.9	19.8	0.2	0.88	0	0.88
PCR 7 WKL	3	3,501	80.0	20.0	0.1	0.87	0	0.87
PCR 8 WE	4	7,164	79.5	19.7	0.8	0.87	0	0.87
PCR 8 WKL	3	7,164	75.8	23.5	0.7	0.87	0.01	0.87
PCR 9 WE	4	3,714	74.0	26.0	0.0	0.85	0	0.85
PCR 9 WKL	3	3,714	74.0	26.0	0.0	0.85	0	0.85
PCR 10 WE	4	3,536	76.5	23.5	0.1	0.88	0.01	0.88
PCR 10 WKL	3	3,536	76.5	23.5	0.1	0.88	0.01	0.88
PCR 11 WE	4	3,537	78.6	21.4	0.0	0.88	0.02	0.88
PCR 11 WKL	3	3,537	78.6	21.4	0.0	0.88	0.02	0.88
PCR 12 WE	4	3,464	80.1	19.8	0.0	0.90	0	0.90
PCR 12 WKL	3	3,464	80.1	19.8	0.0	0.90	0	0.90

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	3	187	95.0	5.0	0.0	0.97	0.02	0.97
PCR 1 WKL	3	187	88.3	11.7	0.0	0.94	0.01	0.94
PCR 2 WE	3	186	88.6	11.4	0.0	0.88	0.03	0.88
PCR 2 WKL	3	186	88.6	11.4	0.0	0.88	0.03	0.88

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	4	158	91.4	8.6	0.0	0.96	0.02	0.96
PCR 1 WKL	3	158	91.4	8.6	0.0	0.96	0.02	0.96
PCR 2 WE	3	158	86.4	13.6	0.0	0.94	0.01	0.94
PCR 2 WKL	3	158	92.2	7.8	0.0	0.96	0.00	0.96

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	3	326	85.7	14.3	0.0	0.89	0.01	0.89
PCR 1 WKL	3	326	85.7	14.3	0.0	0.89	0.01	0.89
PCR 2 WE	3	333	88.4	11.6	0.0	0.93	0.02	0.93
PCR 2 WKL	3	333	88.4	11.6	0.0	0.93	0.02	0.93
PCR 3 WE	3	318	92.5	7.5	0.0	0.95	0	0.95
PCR 3 WKL	3	318	94.5	5.5	0.0	0.96	0.03	0.96
PCR 4 WE	3	332	87.5	12.5	0.0	0.92	0.01	0.92
PCR 4 WKL	3	332	89.1	10.9	0.0	0.86	0.04	0.86

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	3	578	88.1	11.9	0.0	0.94	0.01	0.94
PCR 1 WKL	3	578	86.0	14.0	0.0	0.89	0	0.89
PCR 2 WE	3	577	91.0	9.0	0.0	0.95	0	0.95
PCR 2 WKL	3	577	90.3	9.7	0.0	0.95	0.02	0.95

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
1	2	34,192	89.8	9.8	0.4	0.90	0	0.90
2	2	34,191	87.4	12.4	0.2	0.81	0	0.81
3	2	32,442	76.2	23.3	0.4	0.73	0	0.73
4	2	32,370	81.2	17.5	1.3	0.77	0	0.77
5	2	32,342	77.4	19.0	3.6	0.70	0	0.70
6	2	32,354	79.8	19.4	0.8	0.78	0.01	0.78
7	2	32,422	91.6	7.9	0.5	0.88	0	0.88
8	2	32,418	87.2	11.0	1.7	0.76	0	0.76
9	2	32,464	88.6	11.4	0.0	0.78	0	0.78
10	2	32,499	91.1	6.8	2.0	0.75	0.01	0.75
11	2	32,393	82.4	16.1	1.5	0.87	0	0.87
12	2	34,191	88.6	10.1	1.3	0.89	0	0.89
13	2	34,191	88.3	11.6	0.1	0.88	0	0.88
14	2	34,192	81.1	17.7	1.1	0.69	0	0.69

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
1	2	31,428	89.5	9.7	0.8	0.84	0	0.84
2	2	31,428	92.9	6.8	0.3	0.76	0	0.76
3	2	29,979	89.8	10.2	0.0	0.88	0	0.88
4	2	29,990	88.8	11.2	0.0	0.90	0.01	0.90
5	2	29,967	80.6	19.4	0.0	0.82	0	0.82
6	2	29,992	89.4	10.2	0.4	0.83	0	0.83
7	2	30,010	90.9	8.3	0.8	0.89	0.01	0.89
8	2	29,953	89.6	10.1	0.3	0.82	0	0.82
9	2	29,951	83.9	15.1	1.0	0.80	0	0.80
10	2	29,994	83.7	15.7	0.7	0.79	0	0.79
11	2	30,026	72.9	24.3	2.8	0.69	0.01	0.69
12	2	31,428	81.0	19.0	0.0	0.81	0.01	0.81
13	2	31,428	88.9	11.1	0.0	0.89	0	0.89
14	2	31,428	85.5	14.2	0.3	0.90	0	0.90
15	2	31,427	83.8	15.2	0.9	0.83	0	0.83

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
1	2	19,499	93.5	6.4	0.1	0.75	0	0.75
2	2	19,499	84.2	14.9	0.9	0.87	0	0.87
3	2	18,950	87.9	12.1	0.1	0.89	0	0.89
4	2	18,971	85.8	13.8	0.5	0.66	0	0.66
5	2	18,957	80.4	19.0	0.6	0.79	0	0.79
6	2	18,979	84.3	15.0	0.7	0.76	0.01	0.76
7	2	19,499	87.4	12.2	0.4	0.87	0	0.87
8	2	18,960	80.2	18.5	1.3	0.80	0.01	0.80
9	2	18,953	90.1	9.2	0.7	0.65	0	0.65
10	2	18,958	91.1	8.2	0.8	0.47	0.04	0.47
11	2	18,985	81.0	18.6	0.4	0.82	0	0.82
12	2	18,939	95.7	4.2	0.0	0.94	0	0.94
13	2	19,499	80.6	19.0	0.4	0.71	0	0.71
14	2	19,499	69.6	28.3	2.1	0.65	0	0.65
15	2	19,499	87.9	12.0	0.1	0.83	0	0.83

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
1	2	4,521	79.9	18.7	1.4	0.81	0.02	0.81
2	2	4,500	85.8	13.3	0.9	0.44	0.01	0.44
3	2	4,516	79.0	19.3	1.6	0.75	0.01	0.75
4	2	4,519	84.4	14.5	1.1	0.79	0.02	0.79
5	2	4,519	88.9	10.4	0.8	0.80	0.01	0.80
6	2	4,581	82.8	16.7	0.5	0.86	0.01	0.86
7	2	4,509	83.7	16.2	0.1	0.87	0	0.87
8	2	4,582	78.6	19.3	2.1	0.73	0.01	0.73
9	2	4,556	90.1	9.3	0.6	0.79	0.01	0.79
10	2	4,542	72.4	26.3	1.3	0.73	0.03	0.73
11	2	4,383	82.5	15.8	1.7	0.83	0.03	0.83
12	2	2,677	91.0	9.0	0.0	0.83	0.02	0.83
13	2	4,572	92.4	7.0	0.6	0.94	0	0.94
14	2	4,611	90.0	9.5	0.4	0.90	0	0.90
15	2	4,566	83.9	15.8	0.4	0.79	0	0.79
16	2	4,537	90.3	9.2	0.6	0.87	0.01	0.87
17	2	4,533	81.8	15.9	2.3	0.68	0.01	0.68
18	2	4,519	83.5	16.1	0.4	0.78	0.02	0.78
19	2	4,570	94.9	3.2	1.9	0.77	0.01	0.77
20	2	4,404	80.2	18.1	1.7	0.80	0.03	0.80
21	2	4,537	84.2	15.0	0.9	0.73	0.02	0.73
22	2	4,242	82.2	17.1	0.8	0.86	0.01	0.86
23	2	4,277	89.7	8.8	1.5	0.89	0.01	0.89
24	2	4,508	89.7	9.9	0.4	0.85	0	0.85
25	2	4,510	91.7	8.3	0.1	0.93	0	0.93
26	2	4,131	92.8	7.2	0.1	0.94	0	0.94
27	2	4,552	89.6	9.5	0.9	0.71	0.01	0.71
28	2	4,359	77.7	20.7	1.5	0.78	0	0.78
29	2	4,245	90.3	9.4	0.3	0.91	0	0.91
30	2	4,302	90.3	8.4	1.3	0.86	0.01	0.86
31	2	4,521	89.7	9.9	0.4	0.86	0.01	0.86
32	2	4,277	83.9	15.5	0.6	0.78	0	0.78
33	2	4,530	82.3	17.5	0.2	0.85	0.01	0.85
34	2	4,301	90.9	8.6	0.5	0.89	0	0.89
35	2	4,544	89.3	10.3	0.4	0.88	0	0.88
36	2	4,564	80.5	18.1	1.4	0.83	0.01	0.83
37	2	4,297	90.7	9.2	0.1	0.91	0	0.91
38	2	4,259	77.2	21.3	1.5	0.77	0.01	0.77
39	2	4,592	96.4	3.4	0.2	0.88	0.01	0.88
40	2	4,214	97.8	1.4	0.8	0.96	0	0.96
41	2	4,593	84.3	14.6	1.1	0.74	0.02	0.75
42	2	4,516	89.8	9.8	0.5	0.91	0	0.91

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
43	2	2,642	73.6	26.1	0.3	0.74	0.01	0.74
44	2	4,270	95.1	4.8	0.1	0.93	0	0.93
45	2	4,354	79.9	19.4	0.7	0.63	0.01	0.63
46	2	2,667	90.0	9.3	0.7	0.91	0	0.91
47	2	4,299	80.5	19.0	0.5	0.84	0	0.84
48	2	2,677	91.8	7.8	0.4	0.76	0.02	0.76
49	2	4,299	89.9	9.6	0.6	0.78	0.01	0.78
50	2	2,698	92.7	7.0	0.3	0.81	0.01	0.81
51	2	4,262	98.5	1.5	0.0	0.93	0.01	0.93
52	2	4,519	88.9	10.2	0.9	0.87	0.01	0.87
53	2	4,558	82.4	17.1	0.5	0.83	0.02	0.83
54	2	4,557	89.4	10.6	0.0	0.91	0	0.91
55	2	4,239	77.2	21.8	1.0	0.75	0.01	0.75
56	2	4,395	91.1	6.7	2.2	0.85	0.01	0.85
57	2	4,300	93.0	6.9	0.0	0.87	0	0.87
58	2	4,258	80.6	18.7	0.8	0.78	0.01	0.78
59	2	4,275	86.5	13.1	0.4	0.77	0	0.77
60	2	4,298	78.7	19.7	1.6	0.77	0.01	0.77
61	2	4,281	91.3	8.4	0.3	0.93	0.01	0.93
62	2	4,246	90.1	9.1	0.8	0.92	0.01	0.92
63	2	4,395	90.1	9.6	0.2	0.88	0.01	0.88
64	2	4,360	83.3	16.0	0.6	0.79	0.01	0.79

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
1	2	4,535	84.0	15.3	0.7	0.82	0.01	0.82
2	2	4,540	85.2	14.0	0.8	0.79	0	0.79
3	2	4,571	90.2	8.4	1.5	0.90	0	0.90
4	2	4,568	89.6	9.2	1.2	0.87	0.01	0.87
5	2	4,513	77.5	19.1	3.4	0.78	0.01	0.78
6	2	4,540	90.0	9.1	0.9	0.80	0	0.80
7	2	4,631	90.7	9.3	0.0	0.90	0	0.90
8	2	4,653	93.8	5.9	0.3	0.62	0	0.62
9	2	4,515	81.1	18.3	0.6	0.77	0.01	0.77
10	2	4,520	93.3	6.0	0.7	0.80	0.01	0.80
11	2	4,531	95.1	4.8	0.2	0.63	0.03	0.64
12	2	4,518	89.4	10.3	0.3	0.88	0.01	0.88
13	2	4,577	80.0	17.0	3.0	0.78	0.01	0.78
14	2	4,518	91.5	7.4	1.1	0.62	0.01	0.62
15	2	4,511	92.6	7.0	0.4	0.79	0	0.79
16	2	4,550	90.3	9.0	0.7	0.83	0	0.83
17	2	4,502	84.3	14.9	0.8	0.82	0.01	0.82
18	2	4,676	89.4	9.8	0.7	0.78	0.01	0.78

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
19	2	3,583	86.7	12.9	0.4	0.81	0	0.81
20	2	4,571	81.5	15.9	2.6	0.75	0	0.75
21	2	4,518	72.1	26.2	1.7	0.72	0.03	0.72
22	2	3,513	73.7	24.6	1.7	0.72	0	0.72
23	2	3,583	79.9	19.3	0.8	0.81	0.01	0.81
24	2	4,521	89.7	10.0	0.3	0.92	0	0.92
25	2	4,462	90.0	9.6	0.4	0.90	0	0.90
26	2	4,786	89.5	10.4	0.1	0.92	0	0.92
27	2	4,607	96.5	3.5	0.0	0.89	0.01	0.89
28	2	4,534	90.1	9.9	0.0	0.91	0.01	0.91
29	2	3,583	89.6	10.4	0.0	0.89	0	0.89
30	2	4,518	82.7	15.4	1.9	0.79	0.01	0.79
31	2	4,820	89.6	9.9	0.5	0.89	0.01	0.90
32	2	4,602	90.5	9.0	0.5	0.90	0.01	0.90
33	2	4,556	85.2	14.3	0.5	0.81	0	0.81
34	2	4,478	89.6	10.0	0.4	0.84	0.01	0.84
35	2	4,605	89.5	9.7	0.7	0.87	0	0.87
36	2	4,675	73.6	23.9	2.5	0.72	0	0.72
37	2	4,596	89.6	9.8	0.6	0.89	0.01	0.89
38	2	4,570	93.0	7.0	0.0	0.31	0.02	0.31
39	2	4,601	92.5	7.4	0.1	0.73	0.01	0.73
40	2	4,520	84.3	14.5	1.2	0.74	0.01	0.74
41	2	4,528	89.9	8.3	1.8	0.83	0.02	0.83
42	2	4,539	90.3	8.2	1.5	0.87	0.02	0.87
43	2	3,583	89.7	10	0.3	0.89	0.01	0.89
44	2	4,555	89.1	10.9	0.0	0.80	0.02	0.80
45	2	4,514	92.8	6.5	0.7	0.72	0.02	0.72
46	2	4,549	94.4	5.4	0.1	0.94	0	0.94
47	2	4,671	98.8	1.2	0.0	0.93	0	0.93
48	2	4,599	93.9	5.7	0.4	0.87	0.02	0.87
49	2	4,561	91.4	8.5	0.1	0.93	0.01	0.93
50	2	4,532	83.5	16.4	0.1	0.82	0.01	0.82
51	2	4,531	92.9	7.0	0.1	0.93	0	0.93
52	2	4,601	92.4	7.6	0.0	0.95	0.01	0.95
53	2	4,526	83.7	16.3	0.0	0.81	0.03	0.81
54	2	4,519	95.4	4.6	0.0	0.93	0	0.93
55	2	4,433	91.2	8.8	0.0	0.93	0	0.93
56	2	4,440	83.7	16.3	0.0	0.81	0.06	0.82
57	2	4,619	83.8	15.6	0.6	0.83	0.02	0.83
58	2	4,462	84.7	15.3	0.0	0.84	0.01	0.84
59	2	4,640	90.7	9.2	0.2	0.64	0.01	0.64
60	2	4,639	92.4	7.3	0.3	0.64	0.02	0.64
61	2	3,583	80.3	18.9	0.8	0.78	0.01	0.78

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
62	2	4,719	94.5	5.4	0.1	0.92	0.01	0.92
63	2	4,539	92.3	7.7	0.0	0.81	0.01	0.81
64	2	4,360	83.3	16.0	0.6	0.79	0.01	0.79

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

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
1	2	2,813	89.3	10.7	0.0	0.87	0.01	0.87
2	2	3,746	90.8	8.3	0.9	0.83	0.02	0.83
3	2	2,905	95.9	4.1	0.0	0.74	0.05	0.74
4	2	3,690	94.4	5.6	0.0	0.68	0.02	0.68
5	2	4,604	98.1	1.9	0.1	0.83	0	0.83
6	2	4,537	91.6	8.4	0.0	0.83	0	0.83
7	2	4,527	89.4	9.4	1.2	0.91	0	0.91
8	2	2,805	89.9	9.5	0.6	0.92	0	0.92
9	2	4,513	92.9	6.7	0.4	0.94	0.01	0.94
10	2	2,707	88.3	11.7	0.0	0.91	0.01	0.91
11	2	4,523	88.5	11.4	0.1	0.91	0.01	0.91
12	2	2,909	97.7	2.3	0.0	0.96	0.01	0.96
13	2	2,749	89.1	10.9	0.0	0.79	0.02	0.79
14	2	2,831	90.8	8.5	0.7	0.80	0.01	0.80
15	2	2,739	95.8	4.2	0.0	0.87	0.01	0.87
16	2	2,715	89.5	10.5	0.0	0.78	0.02	0.78
17	2	2,889	95.7	3.4	0.9	0.84	0.01	0.84
18	2	3,729	89.1	10.3	0.6	0.80	0.01	0.80
19	2	3,685	94.5	5.4	0.1	0.60	0.04	0.60
20	2	3,727	93.4	5.9	0.7	0.74	0.03	0.74
21	2	3,103	89.1	10.3	0.6	0.88	0.01	0.88
22	2	2,846	89.1	10.5	0.4	0.81	0.01	0.81
23	2	3,102	88.5	10.8	0.7	0.77	0.01	0.77
24	2	4,640	88.7	10.4	1.0	0.90	0.01	0.90
25	2	4,561	89.4	8.5	2.1	0.87	0	0.87
26	2	2,800	94.1	5.9	0.0	0.76	0.02	0.76
27	2	2,749	94.4	5.6	0.0	0.69	0.04	0.69
28	2	2,893	96.4	3.5	0.1	0.88	0.01	0.88
29	2	4,543	93.2	6.7	0.1	0.93	0	0.93
30	2	3,690	89.1	10.8	0.1	0.69	0	0.69
31	2	4,503	70.9	28.3	0.8	0.68	0.02	0.68
32	2	4,534	79.4	20.5	0.0	0.78	0	0.78
33	2	2,744	86.3	12.6	1.1	0.82	0.03	0.82
34	2	2,757	88.5	11.3	0.2	0.89	0.01	0.89
35	2	4,518	92.1	7.2	0.6	0.82	0.01	0.82
36	2	3,130	86.5	13.4	0.2	0.85	0	0.85
37	2	2,734	98.9	1.0	0.1	0.97	0	0.97

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
38	2	2,803	89.1	10.8	0.0	0.92	0	0.92
39	2	2,810	89.6	9.9	0.5	0.91	0	0.91
40	2	2,822	90.5	9.5	0.0	0.62	0.03	0.62
41	2	2,827	90.4	9.6	0.0	0.62	0.01	0.62
42	2	2,822	89.6	9.5	0.9	0.83	0.01	0.83
43	2	2,796	91.9	8.1	0.0	0.84	0.03	0.84
44	2	2,879	94.9	5.0	0.1	0.86	0.01	0.86
45	2	2,796	89.8	10.2	0.0	0.93	0	0.93
46	2	2,877	85.4	14.6	0.0	0.83	0.04	0.83
47	2	2,862	89.5	10.5	0.0	0.70	0.01	0.70
48	2	2,917	96.7	3.3	0.0	0.81	0.01	0.81
49	2	2,924	88.2	10.7	1.1	0.75	0.01	0.75
50	2	2,885	88.5	11.4	0.0	0.80	0	0.80
51	2	2,875	90.3	9.1	0.7	0.82	0.01	0.82
52	2	2,753	89.5	10.0	0.5	0.91	0	0.91
53	2	2,761	88.5	11.4	0.0	0.77	0.01	0.77
54	2	2,845	94.9	5.1	0.0	0.94	0.01	0.94
55	2	3,102	87.3	12.7	0.0	0.87	0.01	0.87
56	2	2,824	89.7	9.8	0.5	0.87	0.02	0.87
57	2	2,848	89.1	10.7	0.2	0.92	0.01	0.92
58	2	2,859	95.3	4.6	0.1	0.68	0	0.68
59	2	2,779	87.1	12.9	0.0	0.90	0.01	0.90
60	2	4,534	89.8	10.2	0.0	0.87	0.02	0.87
61	2	2,809	82.4	17.6	0.0	0.81	0.01	0.81
62	2	2,794	88.7	11.2	0.1	0.74	0.01	0.74
63	2	4,539	92.3	7.7	0.0	0.81	0.01	0.81
64	2	4,360	83.3	16.0	0.6	0.79	0.01	0.79