2021 RICAS Technical Report

Prepared by Cognia and the

Rhode Island Department of Education

7/25/2022

https://www.ride.ri.gov/

Table of Contents

List of Tables

Preface

The following is the technical documentation for the SY 2020–2021 administration of the Rhode Island Comprehensive Assessment System (RICAS) English language arts (ELA) and mathematics tests. It follows a new format for RICAS technical reports. Because the tests administered in RICAS are the Massachusetts Comprehensive Assessment System (MCAS) ELA and mathematics grades 3–8 tests, much of the information related to their technical quality was traditionally referenced in the RICAS technical report by directing readers to MCAS documentation produced by the Massachusetts Department of Elementary and Secondary Education (DESE). This year, the Rhode Island Department of Education (RIDE) is issuing a stand-alone technical report, which includes any necessary technical information from MCAS directly. MCAS Technical Reports in their entirety are available directly on the DESE website: doe.mass.edu/mcas/tech/?section=techreports.

Because some information in the report is provided by Rhode Island (RIDE) and other information is provided by Massachusetts (DESE), Table A-1 presents an overview of the report's chapters and sections. The source column indicates which state (and, in the case of technical tables, which populations) were used to inform the writing in each section of the report.

From a technical standpoint, the SY 2020–2021 RICAS administration was consistent with previous years' administrations. Test design, development, and administration were all conducted according to standard operating procedure. This standardization allows for individual student score interpretations to hold, as in previous years, demonstrating what individual students know and can do.

One difference from previous administrations is that human-scored items were scored in a distributed instead of center-based model. Great effort was made to ensure that distributed scoring could be done with the same rigor and accuracy as in previous administrations. More detail on distributed scoring and scoring consistency can be found in Chapter 4 of this report.

Another difference was the use of pre-equating. Previously, RICAS tests were post-equated using the Stocking and Lord Procedure. In 2021, DESE, in collaboration with their technical advisory committee (MA-TAC), decided to us a primarily pre-equated solution, with the reasoning that using item parameters fixed prior to disruptions related to COVID-19 would better preserve the meaning and interpretations of student scores. These pre-equated solutions were used whenever possible, with post-equating by the fixed common item parameter method (FCIP) only occurring to manage items not previously calibrated on the next-generation scale. For more detail on the equating solution, consult section 6.2.3 of this report.

The primary differences are not in administration but instead relate to differences in participation rates from previous administrations and the unprecedented changes in learning and instruction due to COVID-19. The entry for Chapter 6 in the 2021 Differences column in Table A-1 is highlighted to warn readers that comparison of some of the data presented in Chapter 6 to previous years is inappropriate, due to unknown differences in the testing population. Specifically, interpretations of statistics that are population dependent, such as item difficulties, correlations with total score, and reliability in 2021 are dependent on a population that is dissimilar to the population tested in 2021. Those differences include not only changes to learning and instruction but also differences in participation from previous administrations. While participation rates are detailed in Appendix B, it is unclear to what extent the missingness in participation is random and these impacts have not been fully studied or quantified. These same cautions also apply to aggregations presented in referenced MCAS and RICAS Appendices. Readers comparing aggregated scores from 2021 to previous administrations should not infer that any change is necessarily the result of a shift in student ability.

Specifically, these data should not be used by readers to make causal claims about COVID-19–related decelerations or fluctuations in learning. Those cautions are listed for the following reasons:

- The complex substantive mechanisms through which the pandemic has created decelerations or fluctuations in measured learning outcomes is unknown;
- Differences in participation rates relative to past years do not allow for direct comparisons of historical trends as the underlying sample compositions may be notable different for some districts; and
- Variations in learning mode (i.e., in-person, remote) across the school year across districts were driven by curricular activities and other compensating factors (e.g., school policies, environment, family circumstances). These variations were not measured / tracked and, consequently, explanatory variables could not be included in the scoring models.

Table A-1 Information Source (RIDE or DESE) by Chapter and Section and differences

Chapter	Section	Description	Data Source	2021 Differences
1	All	Overview	RIDE	None
2		Test Design and Development		
	2.1	Appropriateness	RIDE	
	2.2	Content Standards	RIDE	
	2.3	Performance Standards	RIDE and DESE	None
	2.4	ELA	DESE	
	2.5	Mathematics	DESE	
	2.6	Item and Test Development	DESE	
3	All	Test Administration	RIDE	None
4	All	Scoring	RIDE	Distributed Scoring
5	All	Reporting	RIDE	None
6		Psychometric Quality		**Some comparisons of the data to previous years are inappropriate. Note the description in the preface for more detail.**
	6.1	Classical Item Analyses	RIDE	Unknown differences in population, learning and instruction
	6.2 & 6.3	IRT Linking and Scaling	DESE	Use of a primarily pre-equated solution with some items brought on scale using fixed common item parameterization to maintain pre-COVID individual student score interpretations. See Section 6.2.3
7		Validity		
	7.1	Test Content	DESE	
	7.2	Response Process	DESE	None
	7.3	Internal Structure	DESE	
	7.4	Relationship to Other Variables	DESE	
	7.5	Valid Use of RICAS Data	RIDE	

Chapter 1. Overview

Purposes for administering RICAS include measuring student proficiency relative to standards. Because these standards did not change across administrations, individual student scores can be interpreted in a similar way to previous administrations.

Another stated purpose of RICAS is the use of assessment results for state and federal accountability and reporting. However, these aggregate interpretations cannot be made in the same way as previous years. The disruptions due to COVID-19 and the non-uniform instructional delivery, makes drawing some inferences or comparisons inappropriate. For example, differences in participation rates do not allow for comparison of historical trends. Similarly, the gap in testing, with no RICAS administration in 2020, also confounds historical trends. For these reasons, RIDE sought and received a waiver from accountability from The United State Department of Education, and test scores were not used for accountability purposes in this administration.

1.1 Purpose of This Report

The purpose of this 2021 RICAS Technical Report is to document the technical quality and characteristics of the 2021 RICAS English language arts (ELA) and mathematics tests in grades 3–8, in order to present evidence of the validity, reliability, and fairness of the use of the tests as part of the Rhode Island state assessment program.

Because the RICAS tests administered in Rhode Island are the MCAS ELA and mathematics tests, much of the information related to their technical quality is provided by the MCAS Technical Reports produced by the Massachusetts Department of Elementary and Secondary Education (DESE). That information has been reproduced in this report for the purpose of clarity and DESE, Massachusetts, and MCAS are all referenced in this report. Additionally, MCAS Technical Reports are available directly on the DESE website: doe.mass.edu/mcas/tech/?section=techreports.

This report contains information specific to the administration of the tests in Rhode Island intended to augment the information reproduced from the MCAS Technical Report, to document any differences in the assessment policies and procedures between Rhode Island and Massachusetts, and to provide additional background information about the RICAS program.

The information contained in this report, in conjunction with the technical documentation prepared by Massachusetts, demonstrates that the grades 3–8 MCAS ELA and mathematics tests are technically sound, function well for students in Rhode Island, and are appropriate instruments to assess the performance of Rhode Island students on the state's content standards.

This report is primarily intended for users with a working understanding of psychometrics and educational measurement. It assumes knowledge of measurement concepts such as reliability and validity as well as statistical concepts of correlation and central tendency. For some sections, the reader is presumed to have basic familiarity with advanced topics in measurement and statistics such as item response theory (IRT) and factor analysis.

1.2 Organization of This Report

This report provides information regarding the spring 2021 administration of the 2021 RICAS tests in ELA and mathematics, including a description and results of analyses conducted to provide evidence of the technical quality and characteristics of those tests.

1.2.1 MCAS and RICAS Comparison

The RICAS tests were administered, scored, and processed by Cognia, the state's assessment contractor for the RICAS tests. Cognia is also the Massachusetts assessment contractor for the MCAS tests. Unless noted in this report, all processes and procedures used in administering, processing, scoring, and reporting of the results of the spring 2021 RICAS tests were identical to the corresponding procedures used for the MCAS tests. Table 1-1 provides a summary of the relationship between key aspects of the RICAS and MCAS testing programs.

Table 1-1 Relationship between 2021 RICAS and MCAS Tests on Critical Test Components

Test Component	RICAS and MCAS
Test Content	Identical
Test Design	Identical
Test Administration	Untimed
Mode of Administration	MA administered only one session in grades 3–8; RI administered the full assessment. RI offers Spanish language form in mathematics.
Administration Platform	MA allowed students in grades 3–8 to take tests remotely; RI administered all tests in-person only.
Scoring	
Machine-scored items	Identical
Hand-scored items	Identical
Psychometric Quality	Identical
Reporting	
Scaled scores	Identical
Achievement levels	Identical

Cognia conducted all the analyses described in this report. The analyses described and presented here are consistent with the types of analyses conducted for the MCAS tests.

All analyses are based only on Rhode Island students, unless otherwise specified.

The specific analyses included in this report were identified by the Rhode Island Technical Advisory Committee (RI-TAC) as necessary and useful to provide evidence of the validity, reliability, and fairness of the use of the MCAS tests as the Rhode Island state assessments in ELA and mathematics in grades 3–8.

This information includes the following:

- Chapter 2: Test Design and Development – information related to the MCAS design and development of the tests used for RICAS
- Chapter 3: Test Administration – information related to test administration policies and procedures, including protocols to monitor test security

- Chapter 4: Scoring – information on hand-scoring procedures for short-answer, constructed-response, and essay items, including information on the level of interrater agreement among raters
- Chapter 5: Reporting – detailed information on the type of student-level test scores reported to parents/guardians and a description of the quality assurance procedures used to ensure the accuracy of the reporting of those results
- Chapter 6: Psychometric Quality – a description of and summary results from the set of analyses conducted with Rhode Island students to demonstrate the technical quality and characteristics of the tests (Statistics provided include Classical Item Statistics; Differential Item Functioning; Reliability, including subgroup reliability; and Decision Consistency/Accuracy.) Section 6.1 analyses were conducted using Rhode Island students. This chapter also includes a description of the IRT linking and scaling procedures, analyses and results conducted by Cognia with MCAS and Massachusetts students. Sections 6.2 and 6.3 analyses were also conducted using Massachusetts students.
- Chapter 7: Validity – information related to validity evidence supporting the intended uses and interpretations of RICAS test scores.

Additionally, a set of appendices is provided, containing the following information:

- Appendix A – Accommodations
- Appendix B – Participation Rates
- Appendix C – Interrater Consistency
- Appendix D – Achievement Level Distributions
- Appendix E – Sample Reports
- Appendix F – Item-Level Classical Statistics
- Appendix G – Score Distributions
- Appendix H – Differential Item Functioning Results
- Appendix I – Reliability

1.3 The Rhode Island Comprehensive Assessment System

The RICAS is Rhode Island's state assessment program in ELA and mathematics at grades 3–8, designed to meet the federal requirements of the Every Student Succeeds Act (ESSA). In addition to fulfilling ESSA assessment requirements, the specific purposes of the RICAS tests are (1) to provide information to parents/guardians and students on Rhode Island student achievement on the state's ELA and mathematics content standards, (2) to provide information to support program evaluation and improvement at the school and district level, and (3) to provide academic achievement and growth information used as part of the state's school accountability program to inform parents/guardians and the public about the performance of Rhode Island schools.

Beginning in the 2017–2018 school year, RIDE adopted the MCAS ELA and mathematics tests as its state assessments in ELA and mathematics at grades 3–8. The tests are administered in Rhode Island

under a licensing agreement with Massachusetts DESE and labeled RICAS for their use in Rhode Island. The use of the MCAS tests at grades 3–8 is part of Rhode Island's transition from the use of the Partnership for the Assessment of Readiness for College and Careers (PARCC) tests at grades 3–8 and high school as its state assessments. In high school, the PARCC tests have been replaced by the SAT.

The adoption of the MCAS tests reflects a continuation of Rhode Island's policy to partner with other states to offer a high-quality state assessment. With the increased assessment requirements of the No Child Left Behind Act in 2001, RIDE determined that it would not be feasible to develop and sustain a high-quality assessment program on its own. From 2003–2014, Rhode Island partnered with New Hampshire, Vermont, and Maine in the New England Common Assessment Program (NECAP). With the adoption of the Common Core State Standards (CCSS) and the creation of national assessment consortia, Rhode Island joined PARCC, administering the PARCC tests from 2015–2017.

As Massachusetts and other states left the PARCC consortium, it was no longer clear that PARCC would be able to offer long-term stability in assessment to support the state's improvement efforts. MCAS, in contrast, has been regarded as a model for high-quality and stable state assessment since its inception in 1998. In 2017, Massachusetts developed MCAS tests to fully align with college- and career-ready content standards and established rigorous performance standards consistent with those established by PARCC. With the updated tests and performance standards in place, Rhode Island began administration of the Massachusetts tests in spring 2018.

Chapter 2. Test Design and Development

There were no changes in test design or development for the SY 2020–2021 administration of the RICAS program. The adherence to previous years' blueprints allows for defensible comparisons of where students are relative to grade-level expectations as outlined in the grades 3–8 ELA and mathematics standards despite COVID-related learning disruptions. Chapter 2 is primarily drawn from MCAS technical reporting and relates to the RICAS use of those assessments.

2.1 Appropriateness of Using Massachusetts Standards

Before adopting the MCAS tests as its state assessment, it was necessary to determine the appropriateness of the Massachusetts content and performance standards for use in Rhode Island. To meet the requirements of the Every Student Succeeds Act (ESSA) and provide valid and useful information to Rhode Island parents/guardians, students, and schools, the state assessments must be aligned to the state's content standards. In addition, to support the state's commitment to ensure that Rhode Island's educational system holds high expectations for all students and that Rhode Island graduates are well prepared for postsecondary education, work, and life, the state must establish rigorous performance standards that signal whether students are on track for success in high school and college and career readiness as they progress through elementary and middle school. The following sections describe the steps taken by Rhode Island to make this determination.

2.2 Content Standards

In 2010, Rhode Island adopted the Common Core State Standards (CCSS) as its state content standards in ELA and mathematics. In July 2010, the Massachusetts Board of Elementary and Secondary Education also adopted the CCSS in ELA and mathematics as the core of its PK–12 content standards.

In March 2011, Massachusetts adopted revised Curriculum Frameworks in ELA and mathematics, which are the state's academic content standards. As described at the time by Mitchell Chester, Massachusetts Commissioner of Elementary and Secondary Education, the 2011 Curriculum Framework "merges the Common Core State Standards for Mathematics with additional Massachusetts standards and other features." Rhode Island transitioned to the Rhode Island Core Standards for from the Common Core State Standards (CCSS) on March 9, 2021. The Rhode Island Core Standards mirror the Massachusetts Curriculum Frameworks.

2.3 Performance Standards

In addition to the alignment of the tests to Rhode Island's academic content standards, for the MCAS tests to be appropriate for Rhode Island it was essential that the performance standards established for those tests were consistent with the rigorous performance standards that Rhode Island adopted when it began administering the PARCC tests in 2015.

Massachusetts conducted standard setting activities in August 2017 to establish achievement level cut scores on the new MCAS tests using standardized methods consistent with what is used in the professional field. RIDE staff and technical advisors observed those standard setting procedures and analyzed the results of the standard setting process. Although results of the new tests are reported in terms of four achievement levels, *Not Meeting Expectations*, *Partially Meeting Expectations*, *Meeting Expectations*, and *Exceeding Expectations*, rather than the five levels used to report PARCC results, analyses indicate that the MCAS performance standards are consistent with and as rigorous as the PARCC performance standards previously used in Rhode Island.

Across all grade levels 3–8, results from Rhode Island and Massachusetts suggest that performance at the Meeting Expectations level on the MCAS tests (level 3) is roughly equivalent to performance at the Met Expectations level on the PARCC tests (level 4), in terms of the resulting proportions of students classified above and below those levels.

Cutpoints for grades 3–8 ELA and mathematics RICAS tests were set via standard setting in 2017 by DESE and MCAS for grades 3–8 ELA and mathematics tests (see the *2017 Next-Generation MCAS and MCAS-Alt Technical Report* for the 2017 standard setting report). The standard setting establishes the theta cutpoints used for reporting each year. These theta cuts are presented in Table 2-1. The operational θ-metric cut scores will remain fixed throughout the assessment program unless standards are reset. Also shown in the table are the cutpoints on the reporting score scale.

Table 2-1 Cut Scores on the Theta Metric and Reporting Scale by Content Area and Grade

Content Area	Grade	Theta			Scale Score				
		Cut 1	Cut 2	Cut 3	Min	Cut 1	Cut 2	Cut 3	Max
ELA	3	-1.581	0.011	1.604	440	470	500	530	560
	4	-1.561	0.031	1.623	440	470	500	530	560
	5	-1.659	0.038	1.734	440	470	500	530	560
	6	-1.591	-0.011	1.570	440	470	500	530	560
	7	-1.560	0.011	1.582	440	470	500	530	560
	8	-1.456	0.051	1.559	440	470	500	530	560
Mathematics	3	-1.377	0.027	1.432	440	470	500	530	560
	4	-1.379	0.054	1.487	440	470	500	530	560
	5	-1.551	0.025	1.601	440	470	500	530	560
	6	-1.518	-0.008	1.502	440	470	500	530	560
	7	-1.414	0.031	1.476	440	470	500	530	560
	8	-1.496	-0.008	1.479	440	470	500	530	560

2.4 ELA

2.4.1 ELA Standards

The 2021 RICAS grades 3–8 ELA tests, including all matrix items, were aligned to and measured the following learning standards.

- Anchor Standards for Reading
  - Key Ideas and Details (Standards 1–3)
  - Craft and Structure (Standards 4–6)
  - Integration of Knowledge and Ideas (Standards 7–9)

- Anchor Standards for Language
  - Conventions of Standard English (Standards 1 and 2)
  - Knowledge of Language (Standard 3)
  - Vocabulary Acquisition and Use (Standards 4–6)
- Anchor Standards for Writing
  - Text Types and Purposes (Standards 1–3)
  - Production and Distribution of Writing (Standards 4–6)

The *2017 Massachusetts Curriculum Framework for English Language Arts and Literacy* can be found at www.doe.mass.edu/frameworks/ela/2017-06.pdf.

2.4.2 ELA Item Types

The grades 3–8 ELA tests used several item types, as shown in Table 2-2.

Table 2-2 ELA Item Types and Score Points

Item Type	Possible Raw Score Points	Grade Levels
Multiple-choice (SR)	0 or 1	3–8
Two-part, multiple-choice (SR)	0, 1, or 2	3–8
Technology-enhanced (SR)	0, 1, or 2	3–8
Constructed-response (CR)	0, 1, 2, or 3	3–4
Essay (ES)	0 to 7	3–5
	0 to 8	6–8

SR = selected-response, CR = constructed-response, ES = essay

2.4.3 ELA Passage Types

Passages used in the ELA tests are authentic published passages reviewed by test developers, including DESE test developers, to find passages that possess the characteristics required for use in ELA tests. Passages must

- be of interest to and appropriate for students in the grade being addressed;
- have a clear beginning, middle, and end;
- contain appropriate content;
- support the development of a sufficient number of unique assessment items; and
- be free of bias and sensitivity issues.

Passages ranged in length from approximately 600 to 2500 words per passage set. Word counts were slightly reduced at lower grades. Passage sets consisted of either a single passage or paired/tripled passages. Passages were selected from published works; no passages were specifically written for the RICAS tests.

Passages are categorized into one of two types:

- **Literary passages**—Literary passages represent a variety of genres: poetry, drama, fiction, biographies, memoirs, folktales, fairy tales, myths, legends, narratives, diaries, journal entries, speeches, and essays. Literary passages are not necessarily fictional passages.
- **Informational passages**—Informational passages are reference materials, editorials, encyclopedia articles, and general nonfiction. Informational passages are drawn from a variety of sources, including magazines, newspapers, and books.

In grades 3–8, there is one common form per grade. Each common form included three passage sets, with forms in some grades containing two literary passage sets and one informational passage set, while forms in other grades contained one literary passage set and two informational passage sets. Across the forms, sets may be single, paired, or tripled selections.

The RICAS ELA test is designed to include a set of passages with a balanced representation of male and female characters; races and ethnicities; and urban, suburban, and rural settings. Another important consideration is that passages be of interest to the age group being tested.

The main difference among the passages used for grades 3–8 is their degree of complexity, which results from increasing levels of sophistication in language and concepts, as well as passage length. Test developers use a variety of readability formulas to aid in the selection of passages appropriate at each grade level. In addition, subject-matter experts use their grade-level expertise when participating in passage selection as members of the Assessment Development Committees (ADCs).

2.4.4 ELA Test Design

All items are coded to ELA framework standards. There are no stand-alone items on the tests; all vocabulary, grammar, and mechanics questions are associated with a passage set.

Students read a passage set and answer questions that follow. Question types include selected-response items, constructed-response items (grades 3 and 4 only), and essay items. Please see section 2.4.2 above for additional details on item types. Approximately 20% of the items were technology-enhanced items.

Test Design by Grade

Grades 3–4

The common portion of each test at grades 3 and 4 included three passage sets, and the matrix portion included two passage sets. One of the common passage sets included eleven or twelve 1- or 2-point selected-response items plus one 7-point text-based essay item, one of them included eleven or twelve 1- or 2-point selected-response items and one 3-point constructed-response item, and one of them included six 1- or 2- point selected-response items. Each test contained a total of 44 common points distributed across two testing sessions.

Grade 5

The common portion of each test at grade 5 included three passage sets, and the matrix portion included one passage set. Each passage set included seven or eight 1- or 2-point selected-response items, and one 7-point text-based essay item. The test contained a total of 48 common points distributed across two testing sessions.

Grades 6–8

The common portion of each test at grades 6–8 included three passage sets, and the matrix portion included one passage set. Each passage set included seven or eight 1- or 2-point selected-response

items, and one 8-point text-based essay item. The test contained a total of 51 common points distributed across two testing sessions.

Matrix

For grades 3–8, the matrix portion included one passage set. In grades 3–4, the matrix passage set included eight or nine 1- or 2-point selected-response items, and either two constructed-response items or one essay. In grades 5–8, the matrix passage set included eight or nine 1- or 2-point selected-response items, and one essay item.

Table 2-3 shows the recommended testing times. RICAS tests are untimed; therefore, times shown in the table are approximate.

Table 2-3 ELA Recommended Testing Times, Grades 3–8

Grade	Session 1 Recommended Testing Time (min)	Session 2 Recommended Testing Time (min)	Total Recommended Testing Time (min)
3	120–150	120–150	240–300
4	120–150	120–150	240–300
5	120–150	120–150	240–300
6	120–150	120–150	240–300
7	120–150	120–150	240–300
8	120–150	120–150	240–300

Common and Matrix Item Distribution

The grades 3–8 tests were administered to most students on the computer and to some students with accommodations on a paper form. Tables 2-4 (for the computer-based forms) and 2-5 (for the paper-based forms) list the distribution of common and matrix items in each 2021 ELA test, by grade.

Table 2-4 ELA Distribution of ELA Common and Matrix Items by Grade and Item Type—Computer-based Test (CBT)

Grade	Test	# of Forms	Items per Form							
			Common				Equating/Matrix			
			SR (1 pt.)	SR (2 pt.)	CR	ES	SR (1 pt.)	SR (2 pt.)	CR	ES
3	ELA	1	15	6	1	1	14	2	0	1
4	ELA	1	15	6	1	1	14	2	0	1
5	ELA	1	17	5	0	3	14	2	0	1
6	ELA	1	15	6	0	3	14	2	0	1
7	ELA	1	15	6	0	3	14	2	0	1
8	ELA	1	15	6	0	3	14	2	0	1

Table 2-5 Distribution of ELA Common and Matrix Items by Grade and Item Type—Paper-based Test (PBT)[1]

Grade	Test	# of Forms	Items per Form							
			Common				Equating			
			SR (1 pt.)	SR (2 pt.)	CR	ES	SR (1 pt.)	SR (2 pt.)	CR	ES
3	ELA	1	15	6	1	1	14	2	0	1
4	ELA	1	15	6	1	1	14	2	0	1
5	ELA	1	17	5	0	3	14	2	0	1
6	ELA	1	15	6	0	3	14	2	0	1
7	ELA	1	15	6	0	3	14	2	0	1
8	ELA	1	15	6	0	3	14	2	0	1

[1] *The paper form is derived from Form 1 of the CBT.*

2.4.5 ELA Blueprints

Table 2-6 shows the target and actual percentages of common item points by reporting category. Reporting categories are based on the anchor standards in the 2017 Massachusetts curriculum framework for ELA.

Table 2-6 Target (and Actual) Distribution of ELA Common Item Points by Reporting Category

Reporting Category	% of Points at Each Grade (+/-5%)					
	3	4	5	6	7	8
Language	25 (21)	25 (25)	30(27)	25 (24)	25 (24)	25 (24)
Reading	55 (61)	55 (57)	45 (48)	45 (47)	45 (47)	45 (47)
Writing	20 (18)	20 (18)	25(25)	30 (29)	30 (29)	30 (29)
Total	100	100	100	100	100	100

2.4.6 ELA Cognitive Levels

Each item on the ELA tests is assigned a cognitive level according to the cognitive demand of the item. Cognitive levels are not synonymous with item difficulty. The cognitive level provides information about each item based on the complexity of the mental processing a student must use to answer the item correctly. The three cognitive levels used in ELA tests are described below.

- **Level I (Identify/Recall)**—Level I items require that the student recognize basic information presented in the text. Examples of skills at this level include identifying main ideas/facts/details; recalling and locating details; identifying genre or setting; and identifying definitions, parts of speech, or functions of punctuation. Key words include identify, list, match, recognize, describe, and distinguish.
- **Level II (Infer/Analyze)**—Level II items require that the student understand a given text by making inferences and drawing conclusions related to the text. Examples of skills at this level include understanding the whole text (Big Picture)/generalizing; interpreting, making connections, visualizing, and forming questions; explaining a character's role/motives; determining whether an idea is fact or opinion; filtering important information and key concepts; and determining the meaning of a word in context. Key words include infer, analyze, describe, interpret, determine, conclude, explain, summarize, and classify.
- **Level III (Evaluate/Apply)**—Level III items require that the student understand multiple points of view and be able to project his or her own judgments or perspectives on the text. Examples of skills at this level include understanding another point of view; analyzing/evaluating an author's purpose, style, and message; arguing/defending a point of view with evidence from the text; using reasoning to determine an outcome; applying information from the text; and synthesizing elements of text(s) to create a whole. Key words include critique, evaluate, analyze, predict, agree/disagree, argue/defend, apply, synthesize, judge, compare, and contrast.

Each cognitive level is represented in the ELA tests.

2.4.7 ELA Reference Materials

The use of bilingual word-to-word dictionaries was allowed during both ELA tests only for current and former English language learner (ELL) students. No other reference materials were allowed during the ELA tests.

2.5 MATHEMATICS

2.5.1 Mathematics Standards

The 2021 RICAS grades 3–8 mathematics tests, including all field-test items, were aligned to, and measured the learning standards from the Rhode Island Core Standards.

- Domains for grades 3–5
  - Operations and Algebraic Thinking
  - Number and Operations in Base Ten
  - Number and Operations—Fractions
  - Geometry
  - Measurement and Data
- Domains for grades 6 and 7
  - Ratios and Proportional Relationships
  - The Number System
  - Expressions and Equations
  - Geometry
  - Statistics and Probability
- Domains for grade 8
  - The Number System
  - Expressions and Equations
  - Functions
  - Geometry
  - Statistics and Probability

The 2017 Massachusetts Curriculum Framework for Mathematics can be found at www.doe.mass.edu/frameworks/math/2017-06.pdf.

2.5.2 Mathematics Item Types

The 2021 mathematics tests included several item types, as shown in Table 2-7.

Table 2-7 Mathematics Item Types and Score Points

Item Type	Possible Raw Score Points	Grade Levels
Multiple-choice (SR)	0 or 1	3–8
Multiple-select (SR)	0 or 1	3–8
Technology-enhanced (SA)/(SR)/(CR)	0 or 1 0, 1, or 2	3 4–8
Short-answer (SA)	0 or 1	3–8
Constructed-response (CR)	0, 1, 2, or 3 0, 1, 2, 3, or 4	3 4–8

2.5.3 Mathematics Test Design

Test Design by Grade

Grade 3

The common portion of the grade 3 test included thirty-six 1-point selected-response or short-answer items and four 3-point constructed-response items. The matrix portion included three 1-point selected-response or short-answer items and one 3-point constructed-response item. The test contained a total of 48 common points distributed across two testing sessions.

Grades 4–6

The common portion of the grades 4–6 tests included thirty-four 1-point selected-response or short-answer items, two 2-point selected-response items, and four 4-point constructed-response items. The matrix portion included two 1-point selected-response or short-answer items, one 2-point selected-response or short-answer item, and one 4-point constructed-response item. Each test contained a total of 54 common points distributed across two testing sessions.

Grades 7–8

The common portion of the grades 7–8 tests included thirty-four 1-point selected-response or short-answer items, two 2-point selected-response items, and four 4-point constructed-response items. The matrix portion included two 1-point selected-response or short-answer items, two 2-point selected-response or short-answer items, and two 4-point constructed-response items. Each test contained a total of 54 common points distributed across two testing sessions. Items in session 2 were developed to assess content where the students may need a calculator. These items were either calculator-neutral (calculators are permitted but not required to answer the question) or calculator-active (students are expected to use a calculator to answer the question).

Table 2-8 shows the distribution of common and matrix points on the 2021 mathematics tests, as well as recommended testing times. Since RICAS tests are untimed, the times shown are approximate.

Table 2-8 Mathematics Recommended Testing Times and Common/Matrix Points per Test, Grades 3–8

Grade	# of Sessions	Session 1 Recommended Testing Time (in minutes)	Session 2 Recommended Testing Time (in minutes)	Total Recommended Testing Time (in minutes)	Common Points	Matrix Points
3	2	90	90	180	48	6
4–8	2	90	90	180	54	8–9

The grades 3–8 mathematics tests were administered to most students on the computer and to some students with accommodations on a paper form. Tables 2-9 (for the computer-based forms) and 2-10 (for the paper form) show the distribution of common and matrix item types.

Table 2-9 Distribution of Mathematics Common and Matrix Items by Grade and Item Type—Computer-based Test (CBT)

Grade	# of Forms	Common				Matrix	
		SR/MS SA/TE		CR		SR/MS SA/TE	CR
		(1 pt.)	(2 pt.)	(3 pt.)	(4 pt.)	(1 or 2 pt.)	(3 or 4 pt.)
3	1	36	0	4	0	3	1
4	1	34	2	0	4	3	1
5	1	34	2	0	4	3	1
6	1	34	2	0	4	3	1
7	1	34	2	0	4	4	2
8	1	34	2	0	4	4	2

Table 2-10 Distribution of Mathematics Common and Matrix Items by Grade and Item Type—Paper-based Test (PBT)

Grade	# of Forms	Common				Matrix	
		SR/MS/SA		CR		SR/MS/SA	CR
		(1 pt.)	(2 pt.)	(3 pt.)	(4 pt.)	(1 or 2 pt.)	(3 or 4 pt.)
3	1	36	0	4	0	3	1
4	1	34	2	0	4	3	1
5	1	34	2	0	4	3	1
6	1	34	2	0	4	3	1
7	1	34	2	0	4	4	2
8	1	34	2	0	4	4	2

2.5.4 Mathematics Blueprints

Tables 2-11 through 2-13 show the target and actual percentages of common item points by reporting category. Reporting categories are based on the Rhode Island Core Standards.

Table 2-11 Target (and Actual) Distribution of Mathematics Common Item Points by Reporting Category, Grades 3–5

Domain	% of Points at Each Grade (+/-5%)		
	3	4	5
Operations and Algebraic Thinking	30 (31)	20 (20)	15 (15)
Number and Operations in Base Ten	15 (17)	20 (20)	30 (30)
Number and Operations – Fractions	20 (19)	30 (30)	25 (24)
Geometry	10 (8)	10 (11)	10 (11)
Measurement and Data	25 (25)	20 (19)	20 (20)
Total	**100**	**100**	**100**

Table 2-12 Target (and Actual) Distribution of Mathematics Common Item Points by Reporting Category, Grades 6 and 7

Domain	% of Points at Each Grade (+/-5%)	
	6	7
Ratios and Proportional Relationships	20 (20)	20 (20)
The Number System	20 (20)	20 (19)
Expressions and Equations	30 (30)	25 (26)
Geometry	15 (15)	15 (15)
Statistics and Probability	15 (15)	20 (20)
Total	**100**	**100**

Table 2-13 Target (and Actual) Distribution of Mathematics Common Item Points by Reporting Category, Grade 8

Domain	% of Points (+/-5%)
The Number System and Expressions and Equations	40 (39)
Functions	20 (20)
Geometry	30 (30)
Statistics and Probability	10 (11)
Total	**100**

2.5.5 Mathematics Cognitive Levels

Each item on the mathematics test is assigned a cognitive level according to the cognitive demand of the item. Cognitive levels are not synonymous with difficulty. The cognitive level provides information about each item based on the complexity of the mental processing a student must use to answer the item correctly. The three cognitive levels used in the mathematics tests are described below.

- **Level I (Recall and Recognition)**—Level I items require that the student recall mathematical definitions, notations, simple concepts, and procedures, and apply common, routine procedures or algorithms (that may involve multiple steps) to solve a well-defined problem.
- **Level II (Analysis and Interpretation)**—Level II items require that the student engage in mathematical reasoning beyond simple recall, in a more flexible thought process, and in enhanced organization of thinking skills. These items require a student to make a decision about the approach needed, to represent or model a situation, or to use one or more non-routine procedures to solve a well-defined problem.
- **Level III (Judgment and Synthesis)**—Level III items require that the student perform more abstract reasoning, planning, and evidence-gathering. To answer questions of this cognitive level, a student must engage in reasoning about an open-ended situation with multiple decision points, represent or model unfamiliar mathematical situations, and solve more complex, non-routine, or less well-defined problems.

Cognitive Levels I and II are represented by items in all grades and across item types. Cognitive Level III is best represented by constructed-response items; Cognitive Level III items were included at each grade, whenever possible.

2.5.6 Mathematics Reference Materials

Rulers were provided to students in grades 3–8. Handheld rulers were provided to students taking the paper version of the mathematics test. Students taking the computer-based mathematics test had access to two separate computer-based rulers: a centimeter ruler and a 1/8-inch ruler; students were not permitted to use handheld rulers on the computer-based test.

Reference sheets were provided to students at grades 5–8. These sheets contain information, such as formulas, that students may need to answer certain items.

The second session of the grades 7–8 mathematics tests was a calculator session. All items included in this session were either calculator-neutral (calculators are permitted but not required to answer the question) or calculator-active (students are expected to use a calculator to answer the question). Each student taking the computer-based grade 7 mathematics test had access to a five-function calculator during session 2 of the mathematics test. Each student taking the computer-based grade 8 mathematics test had access to a scientific calculator during session 2. Students taking the paper-based mathematics tests in grades 7–8 had access to comparable handheld calculators.

2.6 Item and Test Development Process

Table 2-14 provides a detailed view of the item and test development process, in chronological order.

Table 2-14 Overview of Item and Test Development Process

Development Step	Detail of the Process
Select reading passages (for ELA only)	Cognia's test developers find potential passages and present them to DESE for initial approval; DESE-approved passages go to Assessment Development Committees (ADCs) composed of experienced educators, and then to a Bias and Sensitivity Committee (BSC) for review and recommendations. ELA items are not developed until passages have been reviewed by an ADC and a BSC. With the ADC and BSC recommendations, DESE makes the final determination as to which passages will be developed and used on a future RICAS test.
Develop items	Cognia's test developers generate items and edit items from subcontractors that are aligned to Massachusetts standards and specifications.
DESE and educator review of items	Cognia sends draft items to DESE test developers for review. DESE test developers review and edit items prior to presenting the items to ADCs. ADCs review items and make recommendations. BSC reviews items and makes recommendations. DESE test developers edit & revise items based on recommendations from ADC & BSC.
Expert review of items	Experts from higher education and practitioners review all field-tested items for content accuracy. Each item is reviewed by at least two independent expert reviewers. Comments and suggested edits are provided to DESE staff for review.
Benchmark constructed-response items and essays	DESE and Cognia test developers meet to determine appropriate benchmark papers for training of scorers of field-tested constructed-response items and essays. Scoring rubrics and notes are reviewed and edited during benchmarking meetings. During the scoring of field-tested items, Cognia contacts DESE test developers with any unforeseen issues.
Item statistics meeting	ADCs review field-test statistics and recommend items for the common-eligible status, for re-field-testing (with edits, for mathematics, since ELA is passage-based), or for rejection. BSC also reviews items and recommends items to become common-eligible or to be rejected.
Test construction	Before test construction, DESE provides target performance-level cut scores to Cognia's test developers. Cognia proposes sets of common items (items that count toward student scores) and matrix items. Matrix items consist of field-test and equating items, which do not count toward student scores. Each common set of items is delivered with proposed cut scores, including test characteristic curves (TCCs) and test information functions (TIFs). DESE test developers and editorial staff review and edit proposed sets of items. Cognia and DESE test developers and editorial staff meet to review edits and changes to tests. Psychometricians are available to provide statistical information for changes to the common form.
Operational test items	Approved common-eligible items become part of the common item set and are used to determine individual student scores.
Released common items	Approximately 50% of common items in grades 3–8 are released to the public, and the remaining items are returned to the common-eligible pools to be used on future RICAS tests. An item description (a statement specifying the content of the item) is released for each common item (both released and non-released).

2.6.1 Item Development and Review

Initial Item Review

As described in the table above, all passages, items, and scoring guides are reviewed by DESE test developers before presentation to the ADCs for review. Passage selection information can be found in section 2.4.3. The DESE test developers evaluate new items for the following characteristics:

- **Alignment**: Are the items aligned to the standards?
- **Content**: Is the content accurate? Does the item elicit a response that shows a depth of understanding of the subject?
- **Contexts**: Are contexts grade-level appropriate? Are they realistic? Are they interesting to students?
- **Grade-level appropriateness**: Are the content, language, and contexts appropriate for the grade level?
- **Creativity**: Does the item demonstrate creativity with regard to approaches to items and contexts?
- **Distractors**: Have the distractors for selected-response items been chosen based on plausible construct-related errors? What are the distractor rationales?
- **Mechanics**: How well are the items written? Are they grammatically correct? Do they follow the conventions of item writing? Is the wording grade-level appropriate and accessible for all students?
- **Technology**: Are the items scored appropriately? Is the item making the best use of the technology? Is there another type of item that is more appropriate?

After initial review, DESE and the contractor's test developers discuss and revise the proposed item sets in preparation for ADC review.

Assessment Development Committee (ADC) and Bias & Sensitivity Committee (BSC) Reviews

ADCs and the BSCs are each composed of approximately 10–12 Massachusetts educators from across the state. Each ADC and BSC meeting is co-facilitated by DESE and Cognia's test developers. There is an ADC for each content area and grade (e.g., ELA grade 3), and there are two BSCs—one for grades 3–7 and one for grades 8 and 10. All ADC and BSC recommendations remain with each item. ADC and BSC members meet several times a year to review new passages and items, and to review data from field-test items. Members review items using Pearson's online platform ABBI. Each participant enters his or her "vote" and recommendations, and the facilitators record the consensus of the committee. DESE takes the recommendations of the ADCs and the BSCs into consideration and makes the final decision to approve items to become field-test eligible.

ADC Passage Review (ELA Only)

ELA ADCs review passages before any corresponding items are written. Committee members consider all the elements noted in section 2.4.3. For example, if a passage is well known or if the passage comes from a book that is widely taught, then the passage is likely to provide an unfair advantage to those students who are familiar with the work. Committee members vote to accept or reject each passage, and the facilitators record the consensus of the group.

For each passage recommended for acceptance, committee members provide suggestions for item development. They also provide recommendations for the presentation of the passage, including

suggestions for the purpose-setting statement, words to be footnoted or redacted, and graphics, illustrations, or photographs to be included with the text.

ADC Item Review

Once DESE test developers have reviewed and edited new items and scoring guides, the items are reviewed by the ADCs. Committees review items for the characteristics noted above. Members vote to accept, accept with edits (members may include suggested edits), or reject each item. The meeting facilitators record the consensus/majority opinion of the group.

BSC Passage and Item Review

After passages and items have been approved by the ADCs, they are also reviewed by a separate BSC. The role of the committee is to identify whether a passage or item contains material that is likely to significantly favor or disadvantage one group of students for reasons that are not educationally relevant. The purpose of the committee's review is to ensure that the ability to answer an item correctly reflects a student's learning, not cultural opportunities or life experiences. Specifically, a passage or item should be flagged by the committee if it is insensitive or disrespectful to a student's ethnic, religious, or cultural background (including disability, socio-economic status, and regional differences). The BSC votes to accept, accept with edits (including suggested edits), or reject (including their reasoning) each passage or item. The meeting facilitators record the consensus of the group.

External Content Expert Item Review

When items are selected to be included on the field-test portion of the RICAS, they are submitted to expert reviewers for their feedback. The task of the expert reviewer is to consider the accuracy of the content of items. Each item is reviewed by two independent expert reviewers. All experts hold a doctoral degree (either in the content they are reviewing or in the field of education) and are affiliated with institutions of higher education in either teaching or research positions. Each expert reviewer has been approved by the DESE. The External Content Experts recommend either accepting or rejecting the item, including their reasoning. Expert reviewers' comments remain with each item.

Editing of Recommended Items

DESE test developers review the recommendations of the ADC, BSC, and expert reviewers and determine whether to revise an item based on the suggested edits. The items are also reviewed and edited by DESE and Cognia editors to ensure adherence to style guidelines in *The Chicago Manual of Style*, *American Heritage Dictionary*, RICAS Style Guidelines, and to sound testing principles. According to these principles, all items should:

- demonstrate correct grammar, punctuation, usage, and spelling;
- be written in a clear, concise style;
- contain unambiguous descriptions of what is required for a student to attain a maximum score;
- be written at a reading level that allows students to demonstrate their knowledge of the subject matter being tested.

Items that pass the reviews listed in this section are approved to be field-tested.

2.6.2 Field-Testing of Items

Only Massachusetts student data are used for field-test analyses. Rhode Island field-test data are not used for item evaluation. Field-tested items appear in the matrix portions of the tests. Each matrix item is typically answered by a minimum of 1,500 students, resulting in enough responses to yield reliable performance data.

Scoring of Field-Tested Items

All field-tested items, except for constructed-response items and essays, are machine-scored. These items include multiple-choice, multiple-select, short-answer, and technology-enhanced items.

All field-tested constructed-response items and essays are hand-scored. To train scorers, DESE works closely with the scoring staff to refine rubrics and scoring notes, and to select benchmark papers that exemplify the score points and variations within each score point. Approximately 1,500 student responses are scored per field-tested constructed-response item or essay. As with machine-scored items, 1,500 student responses are sufficient to provide reliable results. See Chapter 4 for additional information on scorers and scoring.

Data Review of Field-Tested Items

Data Review by DESE

DESE test developers review all item statistics prior to making them available for review by the ADCs and BSCs. An item displaying statistics that indicate it did not perform as expected is closely reviewed and if it is found to be flawed it is rejected from the pool of items. After ADC and BSC reviews of item statistics, DESE test developers make final decisions regarding any recommendations.

Data Review by ADCs

The ADCs meet to review the field-test items with their associated statistics. ADCs review the following item statistics:

- item difficulty (or mean score for polytomous items),
- item discrimination,
- Differential Item Functioning (DIF): female compared with male, African American/Black compared with White, Hispanic or Latino/a compared with White, English language learners (ELL) and former ELL compared with non-ELL,
- distribution of scores across answer options and score points,
- distribution of answer options and score points across quartiles, and
- distribution of unique student responses (for some items).

The ADCs make one of the following recommendations for each field-tested item:

- accept
- edit and field-test again (this recommendation is made for mathematics items only, since ELA items are passage-based)
- reject

Data Review by BSCs

The BSC also reviews the statistics for the field-tested items. The committee reviews only the items that the ADCs have accepted. The BSC pays special attention to items that show DIF when comparing the following subgroups of test takers:

- female compared with male,
- African American/Black compared with White,
- Hispanic or Latino/a compared with White
- English language learners (ELL) and former ELL compared with non-ELL

2.6.3 Item Selection for Operational Test

Cognia's test developers propose a set of previously field-tested or common, non-released items to be used in the common portion of the test. Test developers work closely with psychometricians to ensure that the proposed tests meet the statistical requirements set forth by DESE. In preparation for meeting with the DESE test developers, the Cognia's test developers consider the following criteria in selecting items to propose for the common portion of the test:

- **Content coverage/match to test design and blueprints**. The test designs and blueprints stipulate a specific number of items per item type and per reporting category for each content area. A broad coverage of standards and cognitive skills is expected. The previous year's common test should also be considered, and items should not be duplicated.
- **Item difficulty and complexity**. Item statistics drawn from the data analysis of items are used to ensure similar levels of difficulty and complexity from year to year as well as high-quality psychometric characteristics. Items can be "reused" if they have not been released and not used the previous year. When an item is reused in the common portion of the test, the latest usage statistics accompany that item.
- **"Clueing" items**. Items are reviewed for any information that might "clue" or help the student answer another item.
- **Item types**. A variety of item types, including approximately 20–30% technology-enhanced items, should populate the common slots.

Field-test items are also selected during form construction. Field-test items are drawn from the field-test eligible pools and should mirror the operational test, to the extent needed. If a standard or reporting category is lacking in the common eligible item pool, items should be chosen to fill this need. During assembly of the test forms, the following criteria are considered:

- **Key patterns**. The sequence of keys (correct answers) is reviewed to ensure that the key order appears random.
- **Option balance**. Items are balanced across forms so that each form contains a roughly equivalent number of key options (As, Bs, Cs, and Ds).
- **"Clueing" items**. Items are reviewed for any information that might "clue" or help the student answer another item.
- **Item types**. A variety of item types should populate the matrix slots.

2.6.4 Operational Test Draft Review

The proposed operational test is posted for DESE to review. DESE test developers consider the proposed items, make recommendations for changes, and then meet with Cognia's test developers to construct the final forms of the tests. After form construction meetings, the test forms enter several rounds of review by test developers and editors. Items are checked to ensure that requested changes were made after the test construction meetings, and to ensure that all items are scoring correctly. In addition, items are checked again for any grammatical or "fatal flaw" errors, and these are corrected before the test forms are published.

2.6.5 Special Edition Test Forms

Students with Disabilities

RICAS is accessible to students with disabilities through the universal design of test items, provision of special edition test forms, and the availability of a range of accommodations and accessibility features for students taking the standard tests. To be eligible to receive a special edition test form, a student must have a disability that is documented either in an individualized education program (IEP) or in a 504 plan. All RICAS operational tests and retests were available in the following special editions for students with disabilities:

- **Large-print**—Form 1 of the operational test was translated into a large-print edition. The large-print edition contains all common and matrix items found in Form 1.
- **Braille**—This form included only the common items found in the operational test. If an item indicates bias toward students with visual disabilities (e.g., if it includes a complex graphic that a student taking the Braille test could not reasonably be expected to comprehend as rendered), then simplification of the graphic is considered, with appropriate rewording of the item text, as necessary. If a graphic such as a photograph cannot be rendered in Braille, or if the graphic is not needed for the student to respond to the item, the graphic is replaced with descriptive text or a caption or eliminated altogether. Three-dimensional shapes that are rendered in two dimensions in print are rendered on the Braille test as “front view,” “top view,” and/or “side view,” and are accompanied where necessary by a three-dimensional wooden or plastic manipulative wrapped in a Braille-labeled plastic bag. Modifications to original test items for the Braille version of the test are made only when necessary, as determined by the Braille test subcontractor and DESE staff, and only when they do not provide clues or assistance to the student or change what the item is measuring. When successful modification of an item or graphic is not possible, all or part of the item is omitted, and may be replaced with a similar item.
- **Screen reader**—This accommodation was available only for those students who are blind or have a visual disability. Students who used a screen reader were also given a separate hard-copy Braille edition test to have the appropriate Braille graphics. All answers are entered onscreen, either by the student using a Braille writing device, or by the test administrator.
- **Text-to-speech**—This functionality was embedded in the grades 3–8 CBTs. Students typically use headphones with this format but may also be tested individually in a separate setting to minimize distractions to other students (from hearing what is being read aloud).

Appendix A details other accommodations that did not require a special edition test form and lists accessibility features that were available to all students, such as screen magnification and highlighting. After testing was completed, RIDE received a list with the number of students who participated in the 2021 RICAS with each accommodation, based on information compiled in the Personal Needs Profile in PearsonAccess Next.

Chapter 3. Test Administration

Although the administrations were standardized and the design was not significantly changed from SY 2018–2019 to SY 2020–2021, some changes in administration were allowed to accommodate practical concerns due to COVID-19 disruptions. For example, RIDE extended the state testing window for mathematics. Additional adjustments to uphold health and safety guidance including masking, social distancing, barriers, etc., as well as flexibilities in scheduling were allowed. Details of adjustments to test administration can be found here: https://www.ride.ri.gov/Portals/0/Uploads/Documents/Instruction-and-Assessment-World-Class-Standards/Assessment/RISAP-TC/TestCoordinator-RISAP-Scheduling-Support-SY20-21.pdf

There were no major irregularities reported. Though the overall participation rate was high, some important differences in participation rates and population demographics were observed in the SY 2020–2021 administration. As a result, year-over-year aggregations of student results and comparison of historical trends should only be used when the context of those differences is studied and known.

3.1 Test Administration Schedule

The standard grades 3–8 RICAS tests were administered in two modes, computer-based and paper-based, during two overlapping periods in spring 2021, as shown in Table 3-1.

Table 3-1 Grades 3–8 ELA and Mathematics Test Administration Schedule

Content Area	Complete the Student Registration/ Personal Needs Profile (SR/PNP) Process	Receive Test Administration Materials	Test Administration Windows	Deadline for Return of Materials to Contractor (for PBT Only)
ELA	2/7/22 – 2/18/22	3/16/22	3/28/22 – 4/29/22	5/3/22
Mathematics	2/7/22 – 2/18/22	4/13/22	4/25/22 – 5/27/22 (extended to 6/1/22)	5/31/22

3.2 Security Requirements

Principals were responsible for ensuring that all test administrators complied with the requirements and instructions contained in the Test Coordinator's Manual and Test Administrator's Manuals. In addition, other administrators, educators, and staff within the school were responsible for complying with the same requirements. Schools and school staff who violated the test security requirements were subject to numerous possible sanctions and penalties, including employment consequences, delays in reporting of test results, the invalidation of test results, the removal of school personnel from future RICAS administrations, and possible licensure consequences for licensed educators.

If test content is breached, quick identification and resolution of the breach are critical to the integrity of a testing program. In addition to reports of breaches in the field, the RICAS program used a propriety system by Pearson to perform web monitoring. The system leverages technology tools and human expertise to identify, prioritize, and monitor sites where sensitive test information may be disclosed. The following strategies were used:

- systematically patrolled the Internet, websites, blogs, discussion forums, video archives, social media, document archives, brain dumps, auction sites, and media outlets;

- identified and verified threats to RICAS test security and notified Pearson (who notified RIDE and Cognia, as required);
- worked systematically through the steps necessary to have infringing content removed if a threat was verified; and
- provided summary reporting that included overall and specific threat analysis

Full security requirements, including details about responsibilities of principals and test administrators, examples of testing irregularities, guidance for establishing and following a document tracking system, and lists of approved and unapproved resource materials, can be found in the *Spring 2021 Test Coordinator's Manual (TCM), Grades 3–8 and* the *2021 Test Administrator's Manuals* (TAMs). In spring 2021, there was one TAM for grades 3–8 CBTs and one TAM for grades 3–8 PBTs. The primary delivery mode was computer-based, with paper-based delivery as accommodation only for students with disabilities.

3.3 Participation Requirements

Students in grades 3–8 are expected to participate in RICAS tests for the grade in which they are enrolled and reported to RIDE through the enrollment census.

Participation requirements and guidelines for ELL students and students with significant disabilities are provided in the sections that follow; the participation rates are presented in Appendix B.

See Part III of the *Test Coordinator's Manual* for information about scheduling test administration, including make-up sessions for students who are absent on the day of testing.

3.4 Students Not Tested on Standard Tests

A very small number of students educated with Rhode Island public funds were not required to take the standard RICAS tests. These students were strictly limited to the following categories:

- First-year EL students who enrolled in U.S. schools after April 1, 2020, for whom ELA testing is not required. (First-year EL students must participate in RICAS or Dynamic Learning Maps (DLM) mathematics tests.) See the *RICAS Accessibility and Accommodations Manual, 2021* for details on how EL students participate in spring 2021 RICAS.
- Students with significant cognitive disabilities who are eligible for the alternate assessment, the Dynamic Learning Maps (DLM) Assessment. For more information, refer to the DLM page of the RIDE website: https://www.ride.ri.gov/InstructionAssessment/Assessment/DLMAssessments.aspx.
- Rare and unique situations in which a student is unable to participate in statewide assessments due to a documented, significant, and incapacitating emergency that extends across the entire (or remaining) test window.

More details about test administration policies and participation requirements for non-disabled students, for students with disabilities, for EL students, and for students educated in alternate settings can be found in the *Test Coordinator's Manual*. Data concerning the number of students tested with accommodations are available in Appendix A of this document.

3.4.1 Spanish Edition Test Forms

Spanish-Speaking Students

Spanish editions of the spring grades 3–8 mathematics test were available to any EL student with a low level of English proficiency who was receiving or had received mathematics instruction in Spanish. The Spanish edition of the grades 3–8 mathematics test contained all common and matrix items found in Form 1 of the operational test.

Cognia employed two independent translators to complete the translation of the grades 3–8 mathematics test to Spanish. The translation process was as follows:

- A set of translation rules or parameters was generated, taking the following into consideration: vocabulary, usage, and consistency over the years. These rules were provided to both translators.
- The first translator translated from English to Spanish. The second translator proofed the work of the first translator. Discrepancies between the two translations were resolved by the first translator.
- The Publishing Department reviewed the graphics in Spanish.
- The script that the teacher read when administering the test was also translated into Spanish and was included as Appendix A of the *Test Administrator's Manual*.

The Spanish editions of the grades 3–8 mathematics test were available in both paper and online formats. Human read-aloud in Spanish was also available to students.

3.5 Administration Procedures

It is the test coordinator's responsibility to coordinate the school's RICAS test administration. This coordination responsibility includes the following:

- understanding and enforcing the test security requirements and test administration protocols
- ensuring that students participate in testing according to the requirements in section 3.2 of this report
- coordinating the school's test administration schedule and ensuring that tests are scheduled during the prescribed testing window, and in the prescribed order
- ensuring that accommodations are properly administered and that transcriptions, if required for any accommodation, are properly completed
- completing the Principal's Certification of Proper Test Administration (PCPA) and ensuring the accuracy of information provided on the form
- providing RIDE with the school's correct contact information

More details about test administration procedures, including ordering test materials, scheduling test administration, designating and training qualified test administrators, identifying testing spaces, meeting with students, providing accurate student information, and accounting for and returning test materials, can be found in the *Test Coordinator's Manual*.

The RICAS program is supported by the RICAS Service Center, which includes a toll-free telephone line and email answered by staff members who provide support to schools and districts. The RICAS Service Center operates weekdays from 7:00 a.m. to 5:00 p.m. (Eastern Time), Monday through Friday.

Chapter 4. Human Scoring

There was a significant change to scoring procedures in SY 2020–2021 versus SY 2018–2019. Previously, RICAS human scoring followed a center-based scoring model where scorers worked in an in-person scoring environment. In response to both industry-wide changes and the necessities of limiting in-person work environments due to the COVID-19 pandemic, RICAS 2020–2021 administrations for all grades and contents requiring human scoring shifted to a virtual/synchronous scoring model.

This model maintained the same stringent quality control measures that were applied in the center-based regional scoring model used previously. The similarity of rater training and behavior to previous administrations suggests that scoring (and by extension the results of that scoring) are comparable to previous administrations.

4.1 Interrater Consistency

Interrater consistency statistics are the result of the processes implemented to ensure valid and reliable hand-scoring of items and, as such, provide evidence of scoring stability. Double-blind scoring was one of the processes used to monitor the quality of the hand-scoring of student responses for constructed-response items. For student constructed-response questions in grades 3–8, 10% were randomly selected and scored independently by two different scorers. Results of the double-blind scoring were used during the scoring process to identify scorers who required retraining or other intervention, and they are presented here as evidence of scoring consistency on the RICAS tests.

A third score was required for any score category in which there was not an exact agreement between scorer one and scorer two. A third score was also required as a confirmation score when either scorer one and/or scorer two provided a score of M for Demonstration of Skills and Concepts and Independence or a score of 1 for Level of Complexity.

Summaries of the interrater consistency results are presented in Tables 4-1 for ELA and 4-2 for mathematics by grade. The tables show the number of score categories, the number of included scores, the percent exact agreement, the percent adjacent agreement, the correlation between the first two sets of scores, the percent of responses that required a third score, and linearly-weighted (LW) Kappa as a measure of agreement. Interrater consistency data are available at the item level in Appendix C.

Table 4-1 Summary of Interrater Consistency Statistics Organized across Items by Content Area and Grade—ELA

Content Area	Grade	Number of		Percent		Correlation	% of Third Scores	LW Kappa
		Score Categories	Included Scores	Exact	Adjacent			
ELA	3	4	1,615	73.62	26.25	0.74	0.25	0.608
		5	825	81.58	18.18	0.84	0.24	0.720
	4	4	1,774	75.70	23.51	0.82	1.47	0.726
		5	884	75.11	23.42	0.85	1.70	0.749
	5	4	1,827	72.30	27.37	0.81	0.60	0.695
		5	1,827	72.36	27.26	0.84	0.60	0.712
	6	4	1,763	74.31	25.41	0.86	1.19	0.759
		6	1,763	71.53	27.45	0.87	1.19	0.750
	7	4	1,808	74.23	24.89	0.86	2.05	0.754
		6	1,808	67.04	31.47	0.84	2.05	0.696
	8	4	1,769	72.75	26.79	0.86	1.36	0.743
		6	1,769	68.91	30.02	0.87	1.36	0.725

Note. LW = linearly-weighted

Table 4-2 Summary of Interrater Consistency Statistics Organized across Items by Content Area and Grade—Mathematics

Content Area	Grade	Number of		Percent		Correlation	% of Third Scores	LW Kappa
		Score Categories	Included Scores	Exact	Adjacent			
Mathematics	3	4	3,559	93.87	5.98	0.97	0.14	0.966
	4	5	3,683	88.79	10.64	0.96	0.57	0.802
	5	5	3,721	87.56	11.26	0.96	1.18	0.937
	6	5	3,537	87.42	11.82	0.96	0.76	0.944
	7	5	3,506	91.30	8.19	0.96	0.51	0.942
	8	5	3,449	84.46	14.47	0.94	1.07	0.909

Note. LW = linearly-weighted

Chapter 5. Reporting

One key change to score reporting in the 2020–2021 administration is the absence of SY 2019–2020 assessment results. Another is that there are unknown and likely important differences in participation between the current and prior administrations. Therefore, it is particularly important this year that other information (e.g., opportunity to learn, mode of learning, access to grade level content, attendance, course grades, etc.) be considered when interpreting the data to plan next steps. For example, the complex mechanisms through which the pandemic has created decelerations or fluctuations in learning is unknown. It is inappropriate to use this year's data to compare to previous years in an attempt to attribute changes in learning specifically to pandemic effects. Additionally, because of participation differences, comparing distributional characteristics of 2021 test scores to those from previous years within or across school sites or districts would require additional information that would shed light on the impact of local conditions on participation and learning; such information is not available and therefore not included in this report.

5.1 Reporting of Results

Results on the RICAS were reported in terms of achievement levels that describe student achievement in relation to established state standards. There are four achievement levels for ELA and mathematics for students in grades 3–8: *Not Meeting Expectations*, *Partially Meeting Expectations*, *Meeting Expectations*, and *Exceeding Expectations*. Students were given a separate achievement-level classification in each content area. Reports are generated at the student level. The achievement level distributions are provided in Appendix D.

Parent/Guardian Reports and student results labels were printed and mailed to districts for distribution to parents/guardians and schools. The Parent/Guardian Report is also available to schools in PearsonAccess Next (PAN). Parent/Guardian Report PDFs were run by grade and school and posted online for school, district, and state access.

5.2 Parent/Guardian Report

The Parent/Guardian Report (based on the MCAS report design) was generated for each student eligible to take the RICAS tests. The report is a stand-alone single page (11" x 17") color report that is folded. Two full-color copies of each student's report were printed: one for the parent/guardian and one for the school's records. The report is designed to present parents/guardians with a detailed summary of their child's RICAS performance and to enable comparisons with other students at the school, district, and state levels.

The first page of the Parent/Guardian Report provides student identification information, including student name, grade, date of birth, Student ID (SASID), school name, and district name. The front page also presents general information about the test, website information for parent/guardian resources, and a summary of the student's results for each content area. This summary provides important information for each content area at a glance, including the student's achievement level, scaled score, range of scores, and growth percentile. New in 2021, the front and back pages contain links to Spotlight videos customized to the student's results. Both English and Spanish links were provided to students who were identified by RIDE. The back page of the report is provided by RIDE.

The inside portion of the report contains the achievement level, scaled score, and standard error of the scaled score for each content area tested. If the student does not receive a scaled score, the reason is displayed after “Your Child’s Achievement Level.” Each achievement level has its own distinct color, and that color is used throughout the report to highlight important report elements based on the student’s achievement level and score. These report elements include the student’s earned achievement level, scaled score, the visual scale’s achievement-level title and achievement-level cut scores, and the comparison of the student’s scaled score to the average scaled score at the student’s school, district, and the state levels. If the student received a score previously, their earned scaled score from that year’s test is also displayed along with the current year scaled score for each content area tested. The previous scaled score is displayed in the color corresponding to the achievement level earned that year. If available, up to 3 years of scores including the current year score is displayed in a table.

A student growth percentile (SGP) for each content area tested is displayed with a comparison to the average SGP for the student’s school and district. An SGP describes the student’s learning over time compared to his or her academic peers (peers are other students with similar scores on previous state tests). In 2021, SGP was not reported in grade 4 reports. SGP methodology is described in more detail in Chapter 6. In 2021, SGPs were calculated using a baseline-referenced academic peer group as opposed to previous years when a cohort-referenced peer group was implemented. Utilizing baseline-referenced SGPs for the 2021 RICAS results helped us understand the amount of growth RI students made relative to their pre-pandemic (2019) peers who scored similarly on previous assessments.

The student’s performance in each content area’s reporting categories is also displayed using pictographs and text that indicates the points earned by the student versus the total points possible in that reporting category. For each reporting category, the average number of points earned by students scoring close to 500 is also displayed for comparison purposes. The student’s performance on individual test questions is reported at the bottom of the results page in a simplified item response grid. The grid indicates the points earned and points possible for each test question included on the grid. Essay questions are indicated on the grid. See Appendix E for a sample report.

5.3 Student Results Labels

The other report that is produced for each student is the *Student Results Label*. The labels are printed and shipped to districts. Each label contains the following information for a student: the student’s name, their SASID, grade, date of birth, test date, school code, school name, district name. The student’s results for each subject are also reported. The student’s earned achievement level and scaled score are provided for each subject tested. If the student does not test in one of the subjects, the not tested reason appears on the label. Files are organized by grade, district, and school. Labels are sorted by last name then first name.

5.4 Reporting Business Requirements

To ensure that RICAS results are processed and reported accurately, a document specifying business requirements is prepared before reporting results. The business requirements are adhered to in the processing and analyses of the RICAS test data and in preparation of the reporting results. These rules specify which, if any, student data need to be excluded from school-, district-, and state-level summary computations. At an individual student level, the business requirements document describes how any special cases should be treated for reporting purposes.

5.5 Quality Assurance

Quality assurance measures are implemented throughout the process of analysis and reporting at Cognia. The data processors and data analysts perform routine quality-control checks of their computer programs. When data are handed off to different units within the data team, the sending unit verifies that the data are accurate before handoff. Additionally, when a unit receives a data set, the first step is to verify the accuracy of the data. Once new report designs were approved by RIDE, reports were run using demonstration data to test the application of the decision rules. The populated reports were then approved by RIDE.

Another type of quality assurance measure used at Cognia is parallel processing. One data analyst is responsible for writing all programs required to populate the student-level and aggregate reporting tables for the administration. Each reporting table is assigned to a second data analyst who uses the decision rules to independently program the reporting table. The production and quality-assurance tables are compared; when there is 100% agreement, the tables are released for report generation.

The third aspect of quality control involves procedures to check the accuracy of reported data. Using a sample of schools and districts, the quality assurance group verifies that the reported information is correct. There are two sets of samples selected that may not be mutually exclusive. The first set includes samples that satisfy all the following criteria:

- one-school district
- two-school district
- multi-school district
- private school
- special school (e.g., a charter school)
- small school that does not have enough students to report aggregations
- school with excluded (not tested) students

The second set of samples includes districts or schools that have unique reporting situations that require the implementation of a decision rule. This set is necessary to ensure that each rule is applied correctly.

The quality-assurance group uses a checklist to implement its procedures. Once the checklist is completed, sample reports are circulated for review by psychometric and program management staff. The appropriate sample reports are then sent to RIDE for review and signoff.

Chapter 6. Psychometric Quality

There were no substantial changes to classical item analysis procedures in SY 2020–2021 versus SY 2018–2019. Interpretations of differences using classical item analyses are always difficult, as such statistics are population dependent. Even so, interpretations cannot be made in the same way as previous years. The disruptions due to COVID-19 and the non-uniform instructional delivery makes comparison of aggregated classical test statistics to previous administrations inappropriate for the purposes of quantifying the differences between testing populations. However, it is still appropriate to use classical item statistics to flag items for potential issues in item quality, especially as these issues are further investigated by content experts for additional analysis.

The IRT analyses featured one important difference between the SY 2020–2021 administration and previous years. Instead of post-equated design that identifies and recalibrates items exhibiting parameter drift, a pre-equated design in which previous years' item parameters were held constant was employed whenever prior item parameter estimates were available. In cases where items were not previously calibrated, those items were brought on scale using a fixed common item parameter method (FCIP). This is different than how equating was traditionally conducted in MCAS - and by extension RICAS - in previous administrations. That design did not use FCIP but instead featured a fully post-equated design relying on a set of items held to previous years' parameters and the rest being brought on scale through the Stocking-Lord procedure. In 2021 Drift and dimensionality analyses were conducted to inform thinking about model fit, but the parameters were held constant in order to stabilize the scale and ensure comparability over time. Specifically, these analyses were used to consider whether unknown differences in the testing population post-COVID resulted in a substantial change in the dimensionality of student response patterns or a mean shift in performance.

6.1 Classical Item Analyses

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

Both qualitative and quantitative analyses have been conducted to ensure that 2021 RICAS items meet these standards. This section presents statistical evaluations in four parts: (1) difficulty indices, (2) item-test correlations, (3) DIF statistics, and (4) dimensionality analyses. The item analyses presented here are based on the statewide administration of the RICAS assessments in spring 2021. Note that the information presented in this section is based only on the operational items, since those are the items on which student scores are calculated.

6.1.1 Classical Difficulty and Discrimination Indices

All selected-response and constructed-response items are evaluated in terms of item difficulty according to standard classical test theory (CTT) practices. Difficulty is defined as the average proportion of points achieved on an item and is measured by obtaining the average score on an item and dividing it by the

maximum possible score for the item. Selected-response items are scored dichotomously (correct vs. incorrect), so, for these items, the difficulty index is simply the proportion of students who correctly answered the item. Constructed-response items and essay items are scored polytomously, meaning that a student can achieve scores other than just 0 or 1 (e.g., 0, 1, 2, 3, or 4 for a 4-point constructed-response item). By computing the difficulty index as the average proportion of points achieved, the indices for the different item types are placed on a similar scale, ranging from 0.0 to 1.0 regardless of the item type. Although this index is traditionally described as a measure of difficulty, it is properly interpreted as an easiness index, because larger values indicate easier items. An index of 0.0 indicates that all students earned 0% of the item points, and an index of 1.0 indicates that all students received full credit for the item (i.e., all the item points).

A summary of the item difficulty and item discrimination statistics for each grade and content area combination is presented in Table 6-1. Note that the statistics are presented for all items as well as separately by item type: selected response (SR), constructed response (CR), and essay (ES). The mean difficulty (*p*-value) and discrimination values shown in the table are within generally acceptable and expected ranges.

Table 6-1 Summary of Item Difficulty and Discrimination Statistics by Content Area and Grade

Content Area	Grade	Item Type	Number of Items	Difficulty		Discrimination	
				Mean	Standard Deviation	Mean	Standard Deviation
ELA	3	ALL	31	0.51	0.16	0.43	0.10
		SR	22	0.53	0.13	0.41	0.08
		CR	7	0.57	0.15	0.47	0.12
		ES	1	0.18	0.11	0.56	0.10
	4	ALL	32	0.55	0.14	0.45	0.10
		SR	24	0.59	0.10	0.41	0.07
		CR	6	0.49	0.14	0.54	0.05
		ES	2	0.27	0.08	0.63	0.08
	5	ALL	33	0.58	0.16	0.43	0.12
		SR	24	0.63	0.13	0.38	0.08
		CR	5	0.55	0.07	0.45	0.09
		ES	4	0.30	0.05	0.68	0.01
	6	ALL	34	0.53	0.13	0.43	0.15
		SR	26	0.57	0.10	0.38	0.09
		CR	4	0.44	0.11	0.40	0.10
		ES	4	0.33	0.09	0.75	0.02
	7	ALL	34	0.55	0.14	0.44	0.12
		SR	26	0.60	0.11	0.40	0.07
		CR	4	0.45	0.03	0.45	0.01
		ES	4	0.30	0.07	0.72	0.02
	8	ALL	34	0.57	0.13	0.47	0.13
		SR	26	0.61	0.09	0.41	0.07
		CR	4	0.57	0.08	0.55	0.06
		ES	4	0.33	0.08	0.74	0.04
Mathematics	3	ALL	40	0.46	0.16	0.47	0.14
		SR	17	0.52	0.14	0.41	0.13
		CR	23	0.41	0.16	0.51	0.14
	4	ALL	40	0.46	0.18	0.50	0.12
		SR	11	0.56	0.20	0.41	0.09
		CR	29	0.42	0.16	0.54	0.10

continued

Content Area	Grade	Item Type	Number of Items	Difficulty		Discrimination	
				Mean	Standard Deviation	Mean	Standard Deviation
Mathematics	5	ALL	40	0.42	0.16	0.44	0.15
		SR	18	0.48	0.15	0.36	0.13
		CR	22	0.38	0.16	0.50	0.13
	6	ALL	40	0.36	0.18	0.43	0.14
		SR	14	0.42	0.22	0.34	0.12
		CR	26	0.33	0.14	0.47	0.13
	7	ALL	40	0.35	0.17	0.46	0.15
		SR	17	0.43	0.15	0.35	0.13
		CR	23	0.30	0.16	0.54	0.11
	8	ALL	40	0.38	0.17	0.45	0.16
		SR	16	0.50	0.14	0.33	0.11
		CR	24	0.30	0.15	0.53	0.13

Caution should be exercised when comparing indices across grade levels for the purpose of comparing students in different grade levels and content areas. Differences may be due not only to differences in the item statistics on the test but also may be affected by differences in student abilities and/or differences in the standards and/or curricula taught in each grade. It is reasonable to compare the indices to common benchmarks in the field for the purpose of confirming the items meet industry recognized standards of quality.

Difficulty indices for selected-response items tend to be higher (indicating that students performed better on these items) than the difficulty indices for constructed-response items because selected-response items can be answered correctly by simply identifying rather than providing the correct answer, or by guessing. Similarly, discrimination indices for those constructed-response items with more than two points tend to be larger than those for dichotomous items because of the greater variability of the former (i.e., the partial credit these items allow). The restriction of range (i.e., only two score categories) in dichotomous items tends to make the discrimination indices lower. Note that these patterns are more consistent within item type, so when interpreting classical item statistics, comparisons should be emphasized among items of the same type.

In addition to the item difficulty and discrimination summaries presented above, item-level CTT statistics are provided in Appendix F. On RICAS items, the item difficulty and discrimination indices are within generally acceptable and expected ranges. Very few items were answered correctly at near-chance or near-perfect rates. Similarly, the positive discrimination indices indicate that students who performed well on individual items tended to perform well overall. There are a small number of items with discrimination indices below 0.20, but none were negative. There were only a few items with low discrimination values, or very high or very low item difficulty values included on the 2021 RICAS. These items were included because their statistical values did not negatively impact the quality of the tests, and their inclusion ensured that content specifications were appropriately covered. Item-level score point distributions are provided for constructed-response items in Appendix G; for each item, the percentage of students who received each score point is presented.

6.1.2 Differential Item Functioning

For the RICAS spring 2021 administration, Differential Item Functioning (DIF) analyses were conducted for all subgroups (as defined in the No Child Left Behind Act) for which the sample size was adequate. Six subgroup comparisons were evaluated for DIF:

- male compared with female

- not ELL compared with ELL[1]
- not economically disadvantaged compared with economically disadvantaged
- White compared with African American or Black
- White compared with Hispanic or Latino
- students with disabilities compared with students without disabilities

The tables in Appendix H present the number of items classified as either “low” or “high” DIF, in total and by group favored. The moderate number of items that exhibited low DIF and several that exhibited high DIF were reviewed by content and educational experts to rule out a source of bias prior to being included on the operational tests. For detailed information about how the DIF procedure was employed, please see the *2021 MCAS Next-Generation Technical Report*.

6.1.3 Dimensionality Analysis

The purpose of dimensionality analysis is to investigate whether violation of the assumption of test unidimensionality is statistically detectable and, if so, (a) the degree to which unidimensionality is violated and (b) the nature of the multidimensionality.

The nonparametric IRT-based methods DIMTEST (Stout, 1987; Stout, Froelich, & Gao, 2001) and DETECT (Zhang & Stout, 1999) were applied to operational items for RICAS online test forms administered during the spring 2021 administrations. A total of 12 test forms were analyzed. The data for each grade were split into a training sample and a cross-validation sample. For all grades, there were over 8,900 student examinees per test form in both ELA and mathematics, so every training sample and cross-validation sample had at least 4,450 students. After randomly splitting the data into training and cross-validation samples, DIMTEST was applied to each data set to see if the null hypothesis of unidimensionality would be rejected. DETECT was then applied to each data set for which the DIMTEST null hypothesis was rejected in order to estimate the effect size of the multidimensionality. DETECT values less than 0.2 indicate very weak multidimensionality (or near unidimensionality); values of 0.2 to 0.4, weak to moderate multidimensionality; values of 0.4 to 1.0, moderate to strong multidimensionality; and values greater than 1.0, very strong multidimensionality (Roussos & Ozbek, 2006).

The results of the DIMTEST analyses indicated that the null hypothesis was rejected at a significance level of 0.01 for every data set except for ELA grade 3, which was rejected at a significance level of 0.05. Because strict unidimensionality is an idealization that almost never holds exactly for a given data set, the statistical rejections in the DIMTEST results were not surprising. Indeed, because of the large sample sizes involved in the data sets, DIMTEST would be expected to be sensitive to even quite small violations of unidimensionality.

DETECT was then used to estimate the effect size for the violations of local independence for all the tests. Table 6-2 displays the multidimensionality effect-size estimates from DETECT.

[1] *ELL = English Language Learner (includes current and former English Language Learners).*

Table 6-2 Multidimensionality Effect Sizes by Grade and Content Area

Content Area	Grade	Multidimensionality Effect Size
ELA	3	0.151
	4	0.125
	5	0.213
	6	0.297
	7	0.295
	8	0.188
	Average	**0.212**
Mathematics	3	0.211
	4	0.159
	5	0.159
	6	0.131
	7	0.101
	8	0.177
	Average	**0.156**

The DETECT values indicate weak or very weak multidimensionality for all the 2021 RICAS test forms.

The way in which DETECT divided the tests into clusters was also investigated to determine whether there were any discernable patterns with respect to the selected-response and constructed-response item types. Inspection of the DETECT clusters indicated that selected-response/constructed-response separation generally occurred much more strongly with ELA than with mathematics. Specifically, for the ELA test forms, every grade had one set of clusters dominated by selected-response items and another set of clusters dominated by writing prompt items. On the mathematics test forms, there was less clear evidence of consistent separation of selected-response and constructed-response items.

In summary, for the 2021 dimensionality analyses, the violations of local independence, as evidenced by the DETECT effect sizes, were either very weak or weak in both mathematics and ELA test forms. The patterns with respect to the selected-response and constructed-response items suggested that ELA tended to display more separation than mathematics, however this separation did not result in an effect size that would suggest use of a unidimensional IRT model is inappropriate.

6.2 IRT Linking and Scaling

As reported in Chapter 1, RICAS uses the Massachusetts MCAS ELA and Mathematics tests. The IRT processes used to link and scale RICAS administrations are managed by DESE and Cognia and leveraged by RIDE for Rhode Island Administrations. This section is reprinted from the MCAS technical manual to provide added clarity within this document:

This section describes the procedures used to calibrate, equate, and scale the MCAS and RICAS tests. During these psychometric analyses, a number of quality-control procedures and checks on the processes were conducted. These procedures included

- evaluations of the calibration processes (e.g., checking the number of cycles required for convergence for reasonableness);
- checking item parameters and their standard errors for reasonableness;
- examination of test characteristic curves (TCCs) and test information functions (TIFs) for reasonableness;
- evaluation of model fit test level, item-level and person-level;
- evaluation of equating items (e.g., delta analyses, b-b analyses, beta analyses);

- examination of a-plots and b-plots for reasonableness; and
- evaluation of the scaling results (e.g., comparing look-up tables to the previous year's).

Section 6.2.3 summarizes the equating procedure and results to place the 2021 next-generation RICAS tests on the same scale as the previous year. An equating report, which provided complete documentation of the quality-control procedures and results, was reviewed by the DESE and approved prior to production of the *Spring 2021 MCAS Tests Parent/Guardian Report*s (Cognia Psychometrics and Research Department, *2020–2021 MCAS Equating Report*, unpublished manuscript).

6.2.1 IRT

All RICAS items are calibrated using IRT. IRT uses mathematical models to define a relationship between an unobserved measure of student performance, usually referred to as theta (*θ*), and the probability [*P(θ)*] of getting a dichotomous item correct or of getting a particular score on a polytomous item (Hambleton, Swaminathan, & Rogers, 1991; Hambleton & Swaminathan, 1985). In IRT, it is assumed that all items are independent measures of the same construct (i.e., of the same *θ*). Another way to think of *θ* is as a mathematical representation of the latent trait of interest. Several common IRT models are used to specify the relationship between *θ* and *P(θ)* (van der linden, 2016; Hambleton & van der Linden, 1997; Hambleton & Swaminathan, 1985). The process of determining the mathematical relationship between *θ* and *P(θ)* is called *item calibration*. After items are calibrated, they are defined by a set of parameters that specify a nonlinear, monotonically increasing relationship between *θ* and *P(θ)*. Once the item parameters are known, an estimate of *θ* for each student can be calculated. This estimate, $\hat{\theta}$, is considered to be an estimate of the student's true score or a general representation of student performance. IRT has characteristics that may be preferable to those of raw scores for equating purposes because it specifically models examinee responses at the item level, and also facilitates equating to an IRT-based item pool (Kolen & Brennan, 2014).

For the 2021 RICAS tests, the three-parameter logistic (3PL) model was used for traditional four-option selected-response items, and the two-parameter logistic (2PL) model was used for binary-scored selected-response and technology-enhanced items (Hambleton & van der Linden, 1997; Hambleton, Swaminathan, & Rogers, 1991). The graded-response model (GRM) was used for polytomous items (Nering & Ostini, 2010), including polytomously scored multi-part items, constructed-response items, and essays.

The 3PL model for selected-response items can be defined as:

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

where
U represents the scored response on an item,
i indexes the items,
j indexes students,
α represents item discrimination,
b represents item difficulty,
c is the pseudo guessing parameter,
θ is the student's latent person parameter, and
D is a normalizing constant equal to 1.701.

For the 2PL model, this equation reduces to the following:

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

In the GRM for polytomous items, an item is scored in $k+1$ graded categories that can be viewed as a set of k dichotomies. At each point of dichotomization (i.e., at each threshold), a two-parameter model can be used to model the probability that a student's response falls at or above a particular ordered category, given θ. This implies that a polytomous item with $k+1$ categories can be characterized by k item category threshold curves (ICTCs) of the 2-PL form:

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

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

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

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

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

The GRM is also commonly expressed as:

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

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

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

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

Note that for a dichotomously scored item, $E(U_i|\theta_j) = P_i(\theta_j)$. For more information about item calibration and determination, see Lord and Novick (1968), Hambleton and Swaminathan (1985), or Baker and Kim (2004).

6.2.2 IRT Results

IRT calibration was conducted using flexMIRT 3.03 (Cai, 2012). IRT calibration was conducted for the CBTs in all grades. Because paper test forms are treated as accommodated forms, item parameters for computer-based items were applied to their paper counterparts. The tables in Appendix J of the 2021 MCAS NextGen Technical Report provide the IRT item parameters and associated standard errors of all operational scoring items on the 2021 RICAS tests. MCAS NextGen Technical report Appendix K contains graphs of the TCCs and TIFs, which are defined below.

TCCs display the expected (average) raw score associated with each θ_j value between -4.0 and 4.0. Mathematically, the TCC is computed by summing the ICCs of all items that contribute to the raw score. Using the notation introduced in section 3.6.1, the expected raw score at a given value of θ_j is as follows:

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

where
i indexes the items (and n is the number of items contributing to the raw score),
j indexes students (here, θ_j runs from -4 to 4), and
$E\left(X\middle|\theta_j\right)$ is the expected raw score for a student of ability θ_j.

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

The TIF displays the amount of statistical information that the test provides at each value of θ_j. Information functions depict test precision across the entire latent trait continuum. There is an inverse relationship between the information of a test and its standard error of measurement (SEM). For long tests, the SEM at a given θ_j is approximately equal to the inverse of the square root of the statistical information at θ_j (Hambleton, Swaminathan, & Rogers, 1991), as follows:

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

Compared to the tails, TIFs are often higher near the middle of the θ distribution where most students are located. This is by design. Test items are often selected with middle difficulty levels and high discriminating powers so that test information is maximized for the majority of candidates who are expected to take a test.

The number of cycles required for convergence for each grade and content area during the IRT analysis can be found in Table 6-3. The number of cycles required for convergence fell within acceptable ranges (less than 150) for all tests.

Table 6-3 Number of Cycles Required for Convergence

Content Area	Grade	Computer-based Initial Cycles
ELA	Grade 3	18
	Grade 4	25
	Grade 5	26
	Grade 6	29
	Grade 7	26
	Grade 8	29
Mathematics	Grade 3	57
	Grade 4	37
	Grade 5	28
	Grade 6	38
	Grade 7	42
	Grade 8	51

6.2.3 Equating

The purpose of equating is to ensure that scores obtained from different forms of a test are comparable to one another. Equating may be used if multiple test forms are administered in the same year; or one year's forms may be equated to those used in the previous year. Equating ensures that students are not given an unfair advantage or disadvantage because the test form they took is easier or harder than that taken by other students. See section 3.2 for more information about how the test development process supports successful equating.

It has been the standard practice to use external post-equating for MCAS and RICAS tests. However, considering the potential learning loss during the pandemic, the MCAS Technical Advisory Committee (MA-TAC) had suggested using pre-equating for this year's test to maintain the interpretability of the scale. Compared to post-equating that uses the 2021 data to update the item parameters, pre-equating fixes the item parameters to previously obtained values, such as through field-testing. Pre-equating could potentially better preserve the meaning of the scale in 2021, considering unknown effect of learning loss on testing data.

One complication for implementing the fully pre-equated solution is that some items in the 2021 tests come from the legacy MCAS, and their original parameters were on the legacy MCAS scale rather than the next-generation MCAS scale. A linear transformation had been conducted to transform the legacy item parameters to the next-generation scale, by using a set of items that were administered in both legacy and next-generation MCAS. However, initial pre-equating fit analysis suggested the transformed parameters of the legacy items had poor fit to the data, indicating the linear transformation failed to generate the best estimates for those legacy items.

To reduce the systematic error in the pre-equated parameters for legacy items, a post-equating was conducted by fixing the item parameters for all next-generation items, including both operational and matrix equating items. The fixed common item parameter (FCIP) method was used to estimate the parameters for the legacy items. This method freely calibrates the parameters for those items, while holding the parameters of the other items constant in the calibration.

As it remains unknown how learning loss may have impacted item statistics, the drift analysis for equating items was not conducted or used to decide which items should be excluded from the anchor set. However, the methods of evaluating the suitability of the equating items were still conducted for

exploratory purpose, including the *a*/*a* analysis, the *b*/*b* analysis, and the rescore analysis. The *a*/*a* or *b*/*b* analysis compares the current year's freely estimated IRT discrimination/difficulty parameters with the previous year's values for equating items and flags an item if its standardized distance to the principal axis line is at or above 3 in absolute value. The rescore analysis evaluates the rater drift by having the current year's rater score a sample of constructed responses from previous years and comparing the current year's scores with previous scores. Results from these analyses are included in the equating report.

6.2.4 Reported Scale Scores

Because the θ scale used in IRT calibrations is not understood by most stakeholders, reporting scales were developed for the 2021 RICAS ELA and mathematics tests in grades 3–8. The reporting scales are linear transformations of the underlying θ scale. As the three θ cutpoints from the standard setting have equal intervals, one single linear transformation was sufficient to transform the θ scale from each performance level category on one reporting scale.

Student scores on the RICAS tests are reported in integer values from 440 to 560. Because the same transformation is applied to all achievement-level categories, and the reported scaled scores preserve the interval scale properties (except for the truncated scaled scores at the lower and upper end of the score scale), it is appropriate to calculate means and standard deviations with scaled scores.

By providing information that is more specific about the position of a student's results, scaled scores supplement achievement-level scores. Students' raw scores (i.e., total number of points) on the 2019 next-generation RICAS tests were translated to scaled scores using a data analysis process called *scaling*, which simply converts from one scale to another. In the same way that a given temperature can be expressed on either the Fahrenheit or the Celsius scale, or the same distance can be expressed in either miles or kilometers, student scores on the 2021 next-generation RICAS tests can be expressed in raw or scaled scores.

It is important to note that converting from raw scores to scaled scores does not change students' achievement-level classifications. Given the relative simplicity of raw scores, it is fair to question why scaled scores for the RICAS are reported instead of raw scores. The answer is that scaled scores make the reporting of results consistent. To illustrate, standard setting typically results in different raw cut scores across content areas. The raw cut score between *Partially Meeting Expectations* and *Meeting Expectations* could be, for example, 35 in grade 3 mathematics but 33 in grade 4 mathematics, yet both of these raw scores would be transformed to scaled scores of 500. It is this uniformity across scaled scores that facilitates the understanding of student performance. The psychometric advantage of scaled scores over raw scores comes from their being linear transformations of θ. Since the θ scale is used for equating, scaled scores are comparable from one year to the next. Raw scores are not.

The scaled scores are obtained by a simple translation of ability estimates ($\hat{\theta}$) using the linear relationship between threshold values on the θ metric and their equivalent values on the scaled score metric. Students' ability estimates are obtained by mapping their raw scores through the TCC. Scale scores are calculated using the following linear equation:

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

where
m is the slope and
b is the intercept.

A separate linear transformation is used for each grade and content area combination. Table 6-4 shows the slope and intercept terms used to calculate the scaled scores for each grade and content area. Note that the values in Table 6-4 will not change unless the standards are reset.

Table 6-4 Scale Score Slopes and Intercepts by Content Area and Grade

Content Area	Grade	Slope	Intercept
ELA	3	18.839	499.785
	4	18.846	499.421
	5	17.686	499.335
	6	18.984	500.202
	7	19.098	499.791
	8	19.900	498.981
Mathematics	3	21.357	499.413
	4	20.938	498.869
	5	19.039	499.525
	6	19.870	500.165
	7	20.758	499.353
	8	20.172	500.170

6.3 RICAS RELIABILITY

6.3.1 Reliability and Standard Errors of Measurement

The approach that was implemented to assess the reliability of the 2021 RICAS tests was the *α* coefficient of Cronbach (1951). For details on the calculation of Cronbach's *α* coefficient, please see the *2021 MCAS Next-Generation Technical Report*. Table 6-5 presents descriptive statistics, Cronbach's *α* coefficient, and the raw score standard error of measurement (SEM) for each content area and grade. Statistics are based on operational items from online test forms, which were taken by most of the student examinee population. The reliability estimates range from 0.87 to 0.93, which is a generally acceptable range.

Table 6-5 Raw Score Descriptive Statistics, Cronbach's Alpha, and SEMs by Content Area and Grade

Content Area	Grade	Number of Students	Raw Score			Alpha (α)	SEM
			Maximum	Mean	Standard Deviation		
ELA	3	9,068	44	20.87	8.71	0.89	2.84
	4	9,248	44	22.09	9.28	0.90	2.87
	5	9,448	48	24.52	9.48	0.90	3.04
	6	8,994	50	23.15	10.34	0.90	3.25
	7	9,168	50	23.67	10.44	0.90	3.24
	8	9,109	50	25.24	10.76	0.92	3.10
Mathematics	3	9,065	48	19.89	11.25	0.93	3.02
	4	9,253	54	22.41	12.78	0.94	3.26
	5	9,434	54	20.54	11.90	0.91	3.49
	6	8,933	54	18.68	10.91	0.91	3.34
	7	9,067	54	16.91	10.91	0.92	3.06
	8	8,985	54	18.74	11.14	0.92	3.24

Because of the dependency of the α coefficients on the test-taking population and the test characteristics, precautions need to be taken when making inferences about the quality of one test by comparing its reliability to that of another test from a different grade or content area. To elaborate, reliability coefficients are highly influenced by test-taking population characteristics such as the range of individual differences

in the group (i.e., variability within the population), average ability level of the population that took the exams, test designs, test difficulty, test length, ceiling or floor effect, and influence of guessing. Hence, "the reported reliability coefficient is only applicable to samples similar to that on which it was computed" (Anastasi & Urbina, 1997, p.107). It is reasonable to compare the indices to common benchmarks in the field for the purpose of confirming the tests meet similar industry recognized standards of quality

6.3.2 Reporting Subcategory Reliability

Reliabilities were calculated for the reporting subcategories within the 2021 RICAS content areas. Results and reporting category descriptions are presented in Appendix I. The reliability coefficients for the reporting subcategories range from 0.43 to 0.87, with a median of 0.72 and a standard deviation of 0.12. Because they are based on a subset of items rather than the full test, subcategory reliabilities were typically lower than were overall test score reliabilities, approximately to the degree expected based on the classical test theory (Haertel, 2006), and interpretations should take this into account. Qualitative differences among grades and content areas once again preclude valid inferences about the reliability of the full test score based on statistical comparisons among subtests.

6.3.3 Subgroup Reliability

The reliability coefficients discussed in the previous section were based on the overall population of students who took the 2021 RICAS online forms. Appendix I presents reliabilities for various subgroups of interest for ELA and mathematics, respectively. Cronbach's α coefficients were calculated based only on the members of the subgroup in question in the computations; values are calculated only for subgroups with 10 or more students. The reliability coefficients for subgroups range from 0.71 to 0.94 across the tests, with a median of 0.90 and a standard deviation of 0.032, indicating that reliabilities are generally within a reasonable range.

For several reasons, the subgroup reliability results should be interpreted with caution. Reliabilities are dependent not only on the measurement properties of a test but also on the statistical distribution of the studied subgroup. For example, subgroup sizes may vary considerably, which results in natural variation in reliability coefficients. Alternatively, α, which is a type of correlation coefficient, may be artificially depressed for subgroups with little variability (Draper & Smith, 1998).

6.3.4 Decision Accuracy and Consistency Results

Decision Accuracy and Consistency (DAC) analyses were conducted for all test forms at each performance achievement level. Results of the DAC analyses are provided in Tables 6-4 and 6-5 for the 2021 RICAS tests.

Table 6-6 includes overall accuracy indices with consistency indices displayed in parentheses next to the accuracy values, as well as overall linearly-weighted kappa values. Overall ranges for accuracy (0.79–0.87), consistency (0.71–0.81), and kappa (0.56–0.71) indicate that most students were classified accurately and consistently with respect to measurement error and chance. Accuracy and consistency values conditional on achievement level are also given. For these calculations, the denominator is the proportion of students associated with a given achievement level. For example, the conditional accuracy value is 0.75 for *Not Meeting Expectations* for the grade 3 ELA test. This figure indicates that among the students whose true scores placed them in this classification, 75% would be expected to be in this classification when categorized according to their observed scores. Similarly, a consistency value of 0.61 indicates that 61% of students with observed scores in the *Not Meeting Expectations* level would be expected to score in this classification again if a second, parallel test form was taken.

Because one use of RICAS tests is the placement of student test scores into achievement levels, an important concern is the accuracy and consistency of decisions around achievement level thresholds. In this case, accuracy at the *Partially Meeting Expectations/Meeting Expectations* threshold is critically important, which summarizes the percentage of students who are correctly classified either above or below the particular cutpoint. Table 6-6 provides the accuracy and consistency estimates and false positive and false negative decision rates at each cutpoint for the 2021 RICAS tests. A false positive is the proportion of students whose observed scores were above the cut and whose true scores were below the cut. A false negative is the proportion of students whose observed scores were below the cut and whose true scores were above the cut.

In Table 6-7, the accuracy and consistency indices at the *Partially Meeting Expectations/Meeting Expectations* threshold range from 0.91–0.95 and 0.87–0.93, respectively. The false positive and false negative decision rates at the *Partially Meeting Expectations/Meeting Expectations* threshold both range from 3%–5%. These results indicate that nearly all students were correctly classified with respect to being above or below the *Partially Meeting Expectations/Meeting Expectations* cutpoint.

Table 6-6 Summary of Decision Accuracy (and Consistency) Results by Content Area and Grade—Overall and Conditional on Achievement Level

Content Area	Grade	Overall	Kappa	Conditional on Achievement Level			
				Not Meeting Expectations	Partially Meeting Expectations	Meeting Expectations	Exceeding Expectations
ELA	3	0.80 (0.72)	0.56	0.75 (0.61)	0.82 (0.76)	0.81 (0.74)	0.69 (0.49)
	4	0.82 (0.75)	0.60	0.80 (0.68)	0.83 (0.78)	0.82 (0.77)	0.54 (0.29)
	5	0.82 (0.75)	0.60	0.85 (0.75)	0.84 (0.79)	0.80 (0.73)	0.65 (0.40)
	6	0.79 (0.71)	0.58	0.85 (0.78)	0.78 (0.71)	0.76 (0.67)	0.75 (0.58)
	7	0.81 (0.74)	0.60	0.85 (0.77)	0.81 (0.75)	0.78 (0.70)	0.71 (0.49)
	8	0.83 (0.76)	0.64	0.88 (0.82)	0.83 (0.77)	0.80 (0.73)	0.60 (0.35)
Mathematics	3	0.85 (0.79)	0.68	0.89 (0.85)	0.82 (0.76)	0.83 (0.77)	0.68 (0.45)
	4	0.87 (0.81)	0.71	0.90 (0.85)	0.86 (0.81)	0.84 (0.77)	0.67 (0.43)
	5	0.84 (0.77)	0.63	0.84 (0.76)	0.84 (0.79)	0.83 (0.75)	0.67 (0.40)
	6	0.84 (0.77)	0.63	0.84 (0.77)	0.84 (0.78)	0.83 (0.75)	0.79 (0.59)
	7	0.85 (0.79)	0.67	0.88 (0.81)	0.84 (0.79)	0.85 (0.78)	0.76 (0.53)
	8	0.86 (0.80)	0.68	0.89 (0.83)	0.85 (0.81)	0.84 (0.75)	0.74 (0.45)

Table 6-7 Summary of Decision Accuracy (and Consistency) Results by Content Area and Grade—Conditional on Cutpoint

Content Area	Grade	Not Meeting Expectations / Partially Meeting Expectations			Partially Meeting Expectations / Meeting Expectations			Meeting Expectations / Exceeding Expectations		
		Accuracy (consistency)	False Positive	False Negative	Accuracy (consistency)	False Positive	False Negative	Accuracy (consistency)	False Positive	False Negative
ELA	3	0.93 (0.90)	0.03	0.04	0.91 (0.88)	0.05	0.04	0.96 (0.95)	0.03	0.01
	4	0.93 (0.90)	0.03	0.04	0.92 (0.88)	0.04	0.04	0.98 (0.97)	0.02	0.00
	5	0.94 (0.91)	0.02	0.04	0.91 (0.87)	0.05	0.04	0.97 (0.96)	0.02	0.01
	6	0.91 (0.88)	0.04	0.05	0.92 (0.89)	0.04	0.04	0.96 (0.94)	0.03	0.01
	7	0.92 (0.88)	0.04	0.05	0.92 (0.89)	0.04	0.04	0.97 (0.96)	0.02	0.01
	8	0.93 (0.90)	0.03	0.04	0.92 (0.89)	0.04	0.04	0.98 (0.97)	0.02	0.00
Mathematics	3	0.92 (0.89)	0.04	0.04	0.94 (0.92)	0.03	0.03	0.98 (0.98)	0.01	0.00
	4	0.93 (0.90)	0.03	0.04	0.95 (0.93)	0.03	0.02	0.99 (0.98)	0.01	0.00
	5	0.90 (0.86)	0.05	0.05	0.94 (0.92)	0.03	0.03	0.99 (0.99)	0.01	0.00
	6	0.89 (0.85)	0.05	0.05	0.95 (0.93)	0.03	0.02	0.99 (0.99)	0.01	0.00
	7	0.92 (0.89)	0.03	0.05	0.94 (0.91)	0.04	0.03	1.00 (0.99)	0.00	0.00
	8	0.92 (0.88)	0.04	0.05	0.95 (0.92)	0.03	0.02	1.00 (1.00)	0.00	0.00

The indices above are derived from Livingston and Lewis's (1995) method of estimating DAC. Livingston and Lewis discuss two versions of the accuracy and consistency tables. A standard version performs

calculations for forms parallel to the form taken. An “adjusted” version adjusts the results of one form to match the observed score distribution obtained in the data. The tables use the standard version for two reasons: (1) This “unadjusted” version can be considered a smoothing of the data, thereby decreasing the variability of the results; and (2) for results dealing with the consistency of two parallel forms, the unadjusted tables are symmetrical, indicating that the two parallel forms have the same statistical properties. This second reason is consistent with the notion of forms that are parallel (i.e., it is more intuitive and interpretable for two parallel forms to have the same statistical distribution).

As with other methods of evaluating reliability, DAC statistics that are calculated based on small groups can be expected to be lower than those calculated based on larger groups. For this reason, the values presented in Tables 6-6 and 6-7 should be interpreted with caution. In addition, it is important to remember that it might be inappropriate to compare DAC statistics across grades and content areas.

Chapter 7. Validity

One purpose of this report is to describe the technical and reporting aspects of the RICAS program that support valid score interpretations. According to the *Standards for Educational and Psychological Testing* (AERA et al., 2014), considerations regarding establishment of intended uses and interpretations of test results—and conformance to these uses—are of paramount importance regarding valid score interpretations. These considerations are addressed in this section.

Many sections of this technical report provide evidence of validity, including sections on test design and development, test administration, scoring, scaling and equating, item analysis, reliability, and score reporting. Taken together, these sections provide a comprehensive presentation of validity evidence associated with the RICAS program.

The evidence within each section is built upon the appropriateness of all the other technical aspects of the assessment as documented in this report. Where changes between the SY 2020–2021 administration and previous years were reported, those changes were made primarily to ensure that evidence remained sufficient to preserve validity of interpretation and use and that the descriptions of evidence sources within this chapter continued to support these interpretations and uses.

7.1 Test Content Validity Evidence

Test content validity demonstrates how well the assessment tasks represent the curriculum and standards for each content area and grade level. Content validity is rooted in the item development process, including how the test blueprints and test items align to the curriculum and standards. All items are developed, edited, administered, reviewed, and scored to represent the expectations from the state curriculum frameworks. This process is described further in Chapter 2.

The following are all components of validity evidence based on test content: item alignment with curriculum framework content standards; item bias, sensitivity, and content appropriateness review processes; adherence to the test blueprint; use of multiple item types; use of standardized administration procedures, with accommodated options for participation; and appropriate test administration training. As discussed earlier, all RICAS items are aligned by education stakeholders to specific curriculum framework content standards, and they undergo several rounds of review for content fidelity and appropriateness.

A 2017 content alignment study, conducted by Boston College researchers under the leadership of Michael Russell, found a high degree of content alignment. For mathematics, over 90% of the domains assessed across the grade level tests showed high levels of alignment. For ELA, alignment was also found to be strong across grade levels and domains. When both the items and essay scoring criteria were considered, over 95% of the alignment considerations were deemed adequate. Only two domains, Grade 7 and Grade 8 Reading Informational Text, were identified as candidates for improved alignment. In addition, analyses of the level of agreement among panel members' ratings showed high levels of agreement for the vast majority of ratings following the consensus process. While the study found a few select opportunities to improve alignment, the results from the analyses provide evidence of strong alignment across the vast majority of the tests examined. For further details of this study, please consult the 2017 MCAS Technical Report.

7.2 Response Process Validity Evidence

Response process validity evidence can be gathered via cognitive interviews and/or focus groups with examinees. It is particularly important to collect this type of information prior to introducing a new test or test format, or when introducing new item types to examinees. DESE ensures that evidence of response process validity is collected and reported for all new MCAS item types used in the next-generation assessments.

DESE conducted a 2019 study to determine the readiness of grade 10 students and educators in Massachusetts schools to respond to the next-generation MCAS items. Two standalone field tests were administered to students in every high school in the state. Data from these standalone field tests were then analyzed to determine the following:

- the psychometric properties of the test items and the field tests
- the response time students took to successfully respond to the test

Student response time data were used to filter out the results of students who did not spend sufficient time on their answers. The data from the remaining motivated students were used to examine item discrimination and ensure that new scoring rubrics were keyed correctly. Next-generation test forms were then developed from these sampled results.

7.3 Internal Structure Validity Evidence

Evidence of test validity based on internal structure is presented in great detail in the discussions of item analyses, reliability, and scaling and linking in Chapter 6. Technical characteristics of the internal structure of the assessments are presented in terms of classical item statistics (item difficulty, item-test correlation), DIF analyses, dimensionality analyses, reliability, SEM, and IRT parameters and procedures. In general, item difficulty and discrimination indices were within acceptable and expected ranges. Very few items were answered correctly at near-chance or near-perfect rates. Similarly, the positive discrimination indices indicate that most items were assessing consistent constructs, and students who performed well on individual items tended to perform well overall. See the individual sections for more complete results of the different analyses.

Furthermore, to evaluate whether different reporting categories constitute statistically different dimensions, item-level confirmatory factor analysis (CFA) was conducted to assess the internal structure of the RICAS ELA and mathematics assessments in grades 3–8 from the SY 2017–2018. The CFA model for each test was specified such that the number of factors equaled the number of reporting categories and each item loaded onto the factor that corresponded to the reporting category to which the given item contributed. The results showed very high correlations between different factors, suggesting that there is very little unique variance among the given set of reporting categories. In other words, different reporting categories are essentially measuring the same thing. These results are highly consistent with the unidimensionality results from the DIMTEST and DETECT analyses. Although the CFA analysis suggested unidimensionality among different reporting categories, the high and positive factor loadings do suggest the items provide good measurement for each reporting category. Unidimensionality, meaning items from one reporting category correlate highly to other reporting categories, can be evidence that students have learned different content areas within each subject in an integrated fashion.

7.4 Validity Evidence in Relationship to Other Variables

DESE continues collecting evidence to evaluate the extent to which the next-generation MCAS and RICAS assessments measure "student readiness for the next level" of schooling, such as readiness for the next grade level, or readiness for postsecondary education. In 2019, DESE conducted concurrent validity studies. The first compared student results on the next-generation MCAS tests to course grades and course-taking in middle school and high school. Specifically, the relationships among MCAS results and student course grades in the respective subjects (in ELA and mathematics) showed that MCAS results were more strongly associated with course grades than other covariates tested, including course level, economic disadvantage, being on an IEP, or being an ELL. In mathematics in grade 8, MCAS achievement levels were significantly associated with taking advanced mathematics courses. Convergent validity evidence was also reported between MCAS test portions and subjects.

In 2019, DESE conducted a study examining predictive validity of grade 8 MCAS results on grade 9 course-taking patterns and GPAs. Results from this study will be published as a white paper on the DESE website at www.doe.mass.edu/mcas/tech/.

References

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

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

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

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

Charter, R. A. (1999). Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *Journal of Clinical and Experimental Neuropsychology, 21*(*4*), 559–566(8).

Chicago Manual of Style (16th ed.). (2003). Chicago: University of Chicago Press.

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

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

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

Haertel, E. H. (2006). Reliability. In R.L. Brennan (Ed). *Educational measurement (pp. 65–110)*. Westport, CT: Praeger Publishers.

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: Joint Committee on Testing Practices. Available from fair-testing.pdf Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement 32*, 179–197.

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

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

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

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

Appendices

APPENDIX A

ACCOMMODATIONS

Table A-1. Numbers of Students Tested with and Without Accommodations by Content Area and Grade

Content Area	Grade	Number of Students Tested	
		With Accommodations	Without Accommodations
ELA	3	719	8,349
	4	887	8,361
	5	981	8,467
	6	821	8,173
	7	790	8,378
	8	733	8,376
Mathematics	3	1,675	7,390
	4	1,812	7,441
	5	1,796	7,638
	6	1,096	7,840
	7	1,053	8,014
	8	986	7,999

Table A-2. Numbers of Students Tested with Accommodations by Accommodation Type and Grade—ELA

Description	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Color Contrast	18	49	16	8	6	2
Black on Cream	6	15	3	0	4	1
Black on Light Blue	12	21	11	8	2	0
Black on Light Magenta	0	2	0	0	0	0
White on Black	0	9	1	0	0	1
Yellow on Blue	0	2	0	0	0	0
Dark Gray on Pale Green	0	0	1	0	0	0
Answer Masking	70	118	118	40	34	37
Large Print Test Edition	1	1	0	0	1	0
Screen Reader Edition	0	3	4	4	1	0
Assistive Technology	0	9	6	1	0	0
Braille Test Edition	0	0	0	0	0	0
Human Read Aloud as a Non-Standard Accommodation	21	30	22	23	36	23
Human Signer as a Standard Accommodation	1	4	2	2	7	2
Human Signer as a Non-Standard Accommodation	0	1	0	0	0	1
Text-to-Speech	154	88	126	143	122	103
Human Scribe as a Non-Standard Accommodation	30	26	20	7	5	4
Speech-to-Text as a Non-Standard Accommodation	48	64	68	43	42	43
Typed Responses	0	0	0	0	0	0
Spell-checker	51	39	43	46	55	39
Word Prediction	33	37	25	34	33	17
Graphic Organizer/Reference Sheet	541	670	770	596	575	538
Any Other accommodations	27	33	19	10	17	8
Bilingual Dictionary and Glossary	86	72	58	160	137	140

Table A-3. Numbers of Students Tested with Accommodations by Accommodation Type and Grade—Mathematics

Description	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Color Contrast	12	47	13	10	7	3
Black on Cream	5	15	3	1	5	1
Black on Light Blue	7	19	8	9	2	2
Black on Light Magenta	0	2	0	0	0	0
White on Black	0	9	1	0	0	0
Yellow on Blue	0	2	0	0	0	0
Dark Gray on Pale Green	0	0	1	0	0	0
Answer Masking	58	117	107	43	31	36
Large Print Test Edition	0	0	0	0	1	0
Screen Reader Edition	0	1	1	2	1	1
Assistive Technology	0	2	4	1	0	0
Braille Test Edition	0	0	0	0	0	0
Human Read Aloud as a Standard Accommodation	81	74	50	38	36	24
Human Signer as a Standard Accommodation	1	4	2	2	7	2
Text-to-Speech	1,327	1,406	1,359	705	639	518
Human Scribe as a Standard Accommodation	25	21	18	3	5	4
Speech-to-Text as a Standard Accommodation	59	56	55	34	40	46
Typed Responses	0	0	0	0	0	0
Calculation Device on Non-Calculator Session	82	81	114	126	182	205
Graphic Organizer/Reference Sheet	547	660	712	539	518	507
Any Other accommodations	0	0	0	0	0	0
Spanish	56	73	76	91	75	99
Bilingual Dictionary and Glossary	102	82	61	148	143	121

APPENDIX B

PARTICIPATION RATES

Table B-1. Summary of Participation by Student Subgroup English Language Arts, Grades 3–8

Description	Number Tested	Percent Tested
All Students	55,035	100.00
ELL	9,039	16.42
Economically Disadvantaged	24,480	44.48
African American	4,946	8.99
Asian	1,817	3.30
Hispanic	15,255	27.72
Native American/Alaska Native	397	0.72
White	29,791	54.13
Pacific Islander/Hawaiian	83	0.15
Multiracial	2,725	4.95
Male	28,109	51.07
Female	26,802	48.70
Special Education	8,239	14.97

Table B-2. Summary of Participation by Student Subgroup Mathematics, Grades 3–8

Description	Number Tested	Percent Tested
All Students	54,740	100.00
ELL	9,154	16.72
Economically Disadvantaged	24,291	44.38
African American	4,910	8.97
Asian	1,826	3.34
Hispanic	15,206	27.78
Native American/Alaska Native	391	0.71
White	29,601	54.08
Pacific Islander/Hawaiian	81	0.15
Multiracial	2,704	4.94
Male	28,015	51.18
Female	26,605	48.60
Special Education	8,190	14.96

APPENDIX C

INTERRATER CONSISTENCY

Table C-1. Item-Level Interrater Consistency Statistics—ELA Grade 3

Item Number	Number of Score Categories	Number of Responses Scored Twice	Percent Exact	Percent Adjacent	Correlation	Percent of Third Scores	Kappa
EL625963791	4	790	72.28	27.47	0.72	0.25	0.601
EL735736712#SCORE_TRAIT_Conv	4	825	74.91	25.09	0.76	0.24	0.615
EL735736712#SCORE_TRAIT_Ideadev	5	825	81.58	18.18	0.84	0.24	0.720

Table C-2. Item-Level Interrater Consistency Statistics—ELA Grade 4

Item Number	Number of Score Categories	Number of Responses Scored Twice	Percent Exact	Percent Adjacent	Correlation	Percent of Third Scores	Kappa
EL810046581	4	890	72.70	26.07	0.79	1.24	0.692
EL812949238#SCORE_TRAIT_Conv	4	884	78.73	20.93	0.84	1.70	0.759
EL812949238#SCORE_TRAIT_Ideadev	5	884	75.11	23.42	0.85	1.70	0.749

Table C-3. Item-Level Interrater Consistency Statistics—ELA Grade 5

Item Number	Number of Score Categories	Number of Responses Scored Twice	Percent Exact	Percent Adjacent	Correlation	Percent of Third Scores	Kappa
EL736478825#SCORE_TRAIT_Conv	4	906	69.65	30.02	0.80	0.77	0.676
EL736478825#SCORE_TRAIT_Ideadev	5	906	70.97	28.59	0.84	0.77	0.707
EL806033603#SCORE_TRAIT_Conv	4	921	74.92	24.76	0.82	0.43	0.715
EL806033603#SCORE_TRAIT_Ideadev	5	921	73.72	25.95	0.84	0.43	0.717

Table C-4. Item-Level Interrater Consistency Statistics—ELA Grade 6

Item Number	Number of Score Categories	Number of Responses Scored Twice	Percent Exact	Percent Adjacent	Correlation	Percent of Third Scores	Kappa
EL735440256#SCORE_TRAIT_Conv	4	873	73.42	26.35	0.86	0.57	0.748
EL735440256#SCORE_TRAIT_Ideadev	6	873	73.65	26.00	0.88	0.57	0.767
EL807016586#SCORE_TRAIT_Conv	4	890	75.17	24.49	0.87	1.80	0.771
EL807016586#SCORE_TRAIT_Ideadev	6	890	69.44	28.88	0.86	1.80	0.733

Table C-5. Item-Level Interrater Consistency Statistics—ELA Grade 7

Item Number	Number of Score Categories	Number of Responses Scored Twice	Percent Exact	Percent Adjacent	Correlation	Percent of Third Scores	Kappa
EL807349832#SCORE_TRAIT_Conv	4	902	76.16	22.95	0.87	1.33	0.767
EL807349832#SCORE_TRAIT_Ideadev	6	902	68.96	30.38	0.86	1.33	0.707
EL807456720#SCORE_TRAIT_Conv	4	906	72.30	26.82	0.85	2.76	0.740
EL807456720#SCORE_TRAIT_Ideadev	6	906	65.12	32.56	0.83	2.76	0.685

Table C-6. Item-Level Interrater Consistency Statistics—ELA Grade 8

Item Number	Number of		Percent		Correlation	Percent of Third Scores	Kappa
	Score Categories	Responses Scored Twice	Exact	Adjacent			
EL810463548#SCORE_TRAIT_Conv	4	886	68.85	30.47	0.86	1.58	0.726
EL810463548#SCORE_TRAIT_Ideadev	6	886	67.95	30.93	0.90	1.58	0.751
EL810733917#SCORE_TRAIT_Conv	4	883	76.67	23.10	0.86	1.13	0.760
EL810733917#SCORE_TRAIT_Ideadev	6	883	69.88	29.11	0.84	1.13	0.698

Table C-7. Item-Level Interrater Consistency Statistics—Mathematics Grade 3

Item Number	Number of		Percent		Correlation	Percent of Third Scores	Kappa
	Score Categories	Responses Scored Twice	Exact	Adjacent			
MA261859A	4	877	89.74	9.81	0.95	0.46	0.914
MA286750A_PA	4	1	100.00	0.00	--	0.00	0.000
MA286750A	4	891	96.63	3.37	0.99	0.00	0.971
MA735851787_ES	4	6	100.00	0.00	--	0.00	0.000
MA297399A	4	897	94.87	5.13	0.97	0.00	0.952
MA735851787	4	869	94.02	5.87	0.95	0.12	0.922
MA286750A_ES	4	6	100.00	0.00	1.00	0.00	1.000
MA261859A_ES	4	6	100.00	0.00	1.00	0.00	1.000
MA297399A_ES	4	6	100.00	0.00	1.00	0.00	1.000

Table C-8. Item-Level Interrater Consistency Statistics—Mathematics Grade 4

Item Number	Number of		Percent		Correlation	Percent of Third Scores	Kappa
	Score Categories	Responses Scored Twice	Exact	Adjacent			
MA287484	5	909	93.62	6.38	0.98	0.00	0.952
MA287484_ES	5	8	75.00	25.00	0.80	0.00	0.667
MA716535935	5	915	96.17	3.39	0.98	0.44	0.962
MA800780932	5	919	87.05	12.51	0.96	0.44	0.916
MA800780932_ES	5	8	75.00	25.00	0.49	0.00	0.385
MA801035466	5	909	78.44	20.13	0.91	1.43	0.824
MA801035466_ES	5	8	87.50	12.50	0.75	0.00	0.714
MA716535935_ES	5	7	100.00	0.00	1.00	0.00	1.000

Table C-9. Item-Level Interrater Consistency Statistics—Mathematics Grade 5

Item Number	Number of		Percent		Correlation	Percent of Third Scores	Kappa
	Score Categories	Responses Scored Twice	Exact	Adjacent			
MA624376704_ES	5	8	100.00	0.00	--	0.00	0.000
MA624377498_ES	5	8	100.00	0.00	--	0.00	0.000
MA704359650_PA	5	1	100.00	0.00	--	0.00	0.000
MA624359515	5	919	81.28	16.00	0.94	2.72	0.873
MA624376704	5	911	86.61	12.73	0.94	0.66	0.881
MA624377498	5	927	90.18	9.28	0.97	0.54	0.927
MA704359650	5	931	91.62	7.52	0.97	0.86	0.941
MA624359515_ES	5	8	100.00	0.00	1.00	0.00	1.000
MA704359650_ES	5	8	100.00	0.00	1.00	0.00	1.000

Table C-10. Item-Level Interrater Consistency Statistics—Mathematics Grade 6

Item Number	Number of Score Categories	Number of Responses Scored Twice	Percent Exact	Percent Adjacent	Correlation	Percent of Third Scores	Kappa
MA703249688_ES	5	9	100.00	0.00	--	0.00	0.000
MA713830373_PA	5	1	100.00	0.00	--	0.00	0.000
MA713831396_PA	5	2	100.00	0.00	--	0.00	0.000
MA311694	5	881	81.95	16.80	0.94	1.25	0.872
MA703249688	5	867	86.39	13.03	0.94	0.58	0.887
MA713830373	5	873	91.29	8.48	0.98	0.23	0.944
MA713831396	5	881	89.56	9.42	0.95	1.02	0.906
MA311694_ES	5	8	100.00	0.00	1.00	0.00	1.000
MA713830373_ES	5	8	100.00	0.00	1.00	0.00	1.000
MA713831396_ES	5	7	100.00	0.00	1.00	0.00	1.000

Table C-11. Item-Level Interrater Consistency Statistics—Mathematics Grade 7

Item Number	Number of Score Categories	Number of Responses Scored Twice	Percent Exact	Percent Adjacent	Correlation	Percent of Third Scores	Kappa
MA311144_ES	5	7	100.00	0.00	--	0.00	0.000
MA804701799_ES	5	8	100.00	0.00	--	0.00	0.000
MA295745	5	845	95.62	4.26	0.96	0.12	0.933
MA295758	5	880	93.30	6.59	0.97	0.11	0.934
MA311144	5	876	84.02	15.41	0.93	0.57	0.860
MA804701799	5	874	92.11	6.64	0.96	1.26	0.922
MA295745_ES	5	8	100.00	0.00	1.00	0.00	1.000
MA295758_ES	5	8	100.00	0.00	1.00	0.00	1.000

Table C-12. Item-Level Interrater Consistency Statistics—Mathematics Grade 8

Item Number	Number of Score Categories	Number of Responses Scored Twice	Percent Exact	Percent Adjacent	Correlation	Percent of Third Scores	Kappa
MA301714_ES	5	8	100.00	0.00	--	0.00	0.000
MA311437_ES	5	6	100.00	0.00	--	0.00	0.000
MA713930945_PA	5	1	100.00	0.00	--	0.00	0.000
MA301714	5	854	83.37	15.46	0.93	1.17	0.853
MA311437	5	823	80.19	17.98	0.92	1.82	0.843
MA704855478	5	859	91.50	8.50	0.95	0.00	0.904
MA713930945	5	880	82.05	16.59	0.92	1.36	0.855
MA704855478_ES	5	7	100.00	0.00	1.00	0.00	1.000
MA713930945_ES	5	11	100.00	0.00	1.00	0.00	1.000

APPENDIX D

ACHIEVEMENT LEVEL DISTRIBUTIONS

Table D-1. Cut Scores on the Theta Metric and Reporting Scale by Content Area and Grade

Content Area	Grade	Theta			Scaled Score				
		Cut 1	Cut 2	Cut 3	Min	Cut 1	Cut 2	Cut 3	Max
ELA	3	-1.581	0.0	11	1.604	440	470	500	530
	4	-1.561	0.0	31	1.623	440	470	500	530
	5	-1.659	0.0	38	1.734	440	470	500	530
	6	-1.591	-0.011	1.570	440	470	500	530	560
	7	-1.560	0.0	11	1.582	440	470	500	530
	8	-1.456	0.0	51	1.559	440	470	500	530
Mathematics	3	-1.377	0.0	27	1.432	440	470	500	530
	4	-1.379	0.0	54	1.487	440	470	500	530
	5	-1.551	0.0	25	1.601	440	470	500	530
	6	-1.518	-0.008	1.502	440	470	500	530	560
	7	-1.414	0.0	31	1.476	440	470	500	530
	8	-1.496	-0.008	1.479	440	470	500	530	560

Table D-2. Achievement-Level Distributions by Grade—ELA

Grade	Achievement Level	Percent in Level	
		2019	2021
3	Not Meeting Expectations	11.55	13.76
	Partially Meeting Expectations	40.55	45.88
	Meeting Expectations	40.07	35.72
	Exceeding Expectations	7.83	4.64
4	Not Meeting Expectations	14.24	16.24
	Partially Meeting Expectations	48.52	48.30
	Meeting Expectations	33.60	32.83
	Exceeding Expectations	3.64	2.63
5	Not Meeting Expectations	12.58	18.36
	Partially Meeting Expectations	48.33	48.26
	Meeting Expectations	35.35	30.00
	Exceeding Expectations	3.75	3.38
6	Not Meeting Expectations	20.95	28.21
	Partially Meeting Expectations	39.94	39.23
	Meeting Expectations	32.64	26.04
	Exceeding Expectations	6.48	6.53
7	Not Meeting Expectations	22.94	26.48
	Partially Meeting Expectations	45.47	44.83
	Meeting Expectations	27.36	25.10
	Exceeding Expectations	4.23	3.59
8	Not Meeting Expectations	23.21	26.89
	Partially Meeting Expectations	40.48	44.33
	Meeting Expectations	30.48	25.60
	Exceeding Expectations	5.82	3.18

Table D-3. Achievement-Level Distributions by Grade—Mathematics

Grade	Achievement Level	Percent in Level	
		2019	2021
3	Not Meeting Expectations	20.49	35.38
	Partially Meeting Expectations	43.40	39.55
	Meeting Expectations	31.12	23.04
	Exceeding Expectations	4.99	2.03
4	Not Meeting Expectations	20.34	33.89
	Partially Meeting Expectations	47.10	45.28
	Meeting Expectations	29.18	19.07
	Exceeding Expectations	3.37	1.75
5	Not Meeting Expectations	18.68	28.70
	Partially Meeting Expectations	51.19	51.03
	Meeting Expectations	28.22	19.06
	Exceeding Expectations	1.91	1.21
6	Not Meeting Expectations	19.03	32.07
	Partially Meeting Expectations	53.02	50.15
	Meeting Expectations	25.31	16.46
	Exceeding Expectations	2.63	1.32
7	Not Meeting Expectations	22.65	30.80
	Partially Meeting Expectations	49.69	48.85
	Meeting Expectations	24.71	18.51
	Exceeding Expectations	2.95	1.84
8	Not Meeting Expectations	24.25	36.57
	Partially Meeting Expectations	51.28	47.38
	Meeting Expectations	21.91	14.91
	Exceeding Expectations	2.55	1.14

APPENDIX E

SAMPLE REPORTS

Rhode Island Education

Accelerating learning by providing the academic, social-emotional, and wrap-around services that our school communities need to ensure a high-quality education for **every single student** in Rhode Island.

Where We Are

The Rhode Island Department of Education (RIDE) launched the Learning, Equity & Accelerated Pathways Task Force, asking state and local leaders to join our efforts to get a better understanding of what support and resources our school communities need to accelerate learning for all students and address the educational inequities statewide.

Over two months, RIDE and the LEAP task force engaged in an evidence-based process relying on data and the knowledge of national education experts.

Through a report, the task force provided Commissioner Infante-Green with recommendations that outlined enabling conditions to define what must be met across systems to be successful and absolute priorities that center the needs of Rhode Island students who are multilingual, differently-abled, or who live in the urban core through an explicitly anti-racist, equity-focused lens.

Where We're Going

By working with LEAs to ensure that every precaution was taken to keep the 2021 RICAS testing safe, teachers supported, and parents well informed about the importance and benefits for their child, we now have a snapshot of how kids are doing in English Language Arts (ELA) and mathematics. State and local leaders are now using this information to better understand what is needed to best support our school communities. This information will also guide how the federal government provides funding to schools to help your child with learning.

Join us-together-we can improve education!
Go to the following link(s) to view a personalized video about your child's results:
English: https://prod.spotlight-education.com/p-Mjkx4xbDIWvttEn

Scan the QR code to access
important information and resources for your family

Paving the Way with Absolute Priorities

The following LEAP absolute priorities accompanied by the 2021 RICAS assessment data will help accelerate student learning and move our pre-kindergarten through grade twelve system forward.

- Energize our school communities by launching a statewide back-to-school campaign
- \+ Ensure all students have access to high-quality & personalized support through extended learning, before/after school partnerships, & summer learning
- \+ Universally screen all students and align resources to need
- \+ Improve and support student transitions across grades and systems
- \+ Close the digital divide

RIDE Rhode Island Department of Education

Learn more at www.RIDE.ri.gov

Spring 2021 RICAS Individual Student Report

RIDE Rhode Island Department of Education

Name:	**District:**
SASID:	**School:**
Date of Birth:	**Grade:** 5

This report provides your child's results from the 2021 Rhode Island Comprehensive Assessment System (RICAS) tests in English Language Arts (reading and writing) and mathematics.

The COVID-19 pandemic brought new challenges to our schools, and parents, teachers, and administrators worked together over the past year to address and overcome these challenges. When reviewing your child's results from this assessment, keep in mind that your child's performance may have been influenced by disruptions due to the COVID-19 pandemic. The pandemic may also have influenced the performance of your child's school, district, and the state. In alignment with the work of the Learning, Equity & Accelerated Pathways (LEAP) Task Force (https://www.ride.ri.gov/InsideRIDE/AdditionalInformation/LEAPTaskForce.aspx), RIDE has remained committed to rebuilding Rhode Island's educational system post-pandemic, helping students get back up to speed, and offering greater access to enriching learning opportunities.

We thank you for your participation in these tests which helped guide this critical work to improve outcomes for students. While it is important to acknowledge the challenges of this past year, we must now focus on understanding your child's understanding of ELA and mathematics knowledge and skills. We hope this report can help inform and empower you as you advocate for your child. You know your child best.

Join us-together-we can improve education!
Go to the following link(s) to view a personalized video about your child's results:
English: https://prod.spotlight-education.com/p-Mjkx4xbDIWvttEn

For each subject, the report shows:
- Your child's score between 440 and 560 and their achievement level
- How your child performed in reading and mathematics based on the test reporting categories
- A growth score that shows how your child performed compared to other students who scored similarly

Your Child's Overall Results

English Language Arts

Achievement Level
Meeting Expectations

Score
518
(Score range: 440-560)

Growth Percentile
60

Details on page 2

Mathematics

Achievement Level
Meeting Expectations

Score
508
(Score range: 440-560)

Growth Percentile
22

Details on page 3

Did you know that establishing family routines can help your child succeed?
Make a habit of setting up designated times for homework, reading, mealtimes, family conversations, bedtime, and leaving for school each day.

What do I do next?
After reviewing this report, it is critical that you attend family-teacher conferences **and discuss with your child's teachers your questions and concerns**. Don't be afraid to speak up. Children whose families stress the value of education are more likely to find it important, as well.

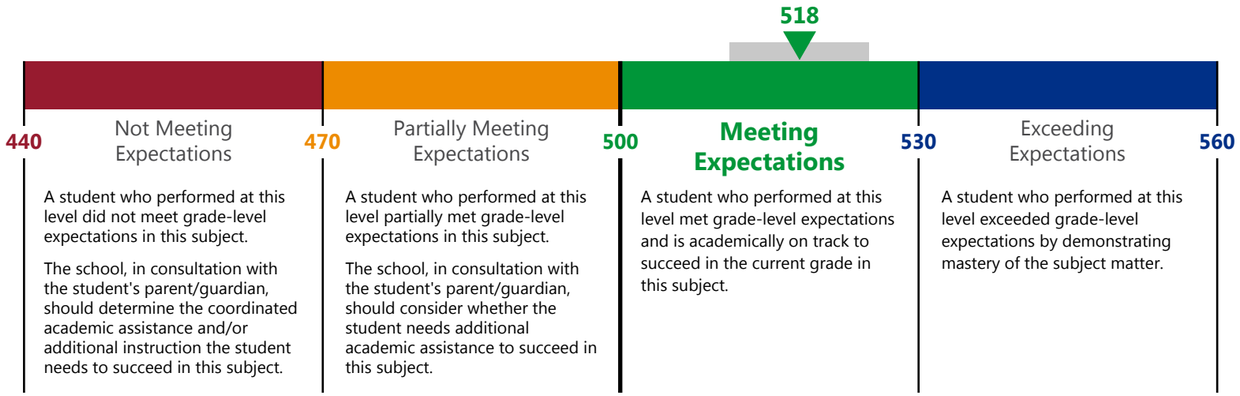
How can I support my child's education?
- School attendance matters, *every single day*. Missing just two days of school a month is chronically absent, so make it a priority to get your child to school on time daily.
- Establish daily reading routines, let your child see *you* read, and encourage your child to read for fun all year long.
- Get involved and stay connected to your child's school, however and whenever you can.
- Share your voice! Help improve your child's school by participating in SurveyWorks every year.
- Start a conversation. Ask questions. Talk to your child about what they're learning, and show an interest in the subjects that excite them.

Name:
SASID:

Grade 5
Spring 2021
Computer-based test

English Language Arts

Your Child's Achievement Level:	**Meeting Expectations**
Your Child's Score:	**518**



Not Meeting Expectations
A student who performed at this level did not meet grade-level expectations in this subject.

The school, in consultation with the student's parent/guardian, should determine the coordinated academic assistance and/or additional instruction the student needs to succeed in this subject.

Partially Meeting Expectations
A student who performed at this level partially met grade-level expectations in this subject.

The school, in consultation with the student's parent/guardian, should consider whether the student needs additional academic assistance to succeed in this subject.

Meeting Expectations
A student who performed at this level met grade-level expectations and is academically on track to succeed in the current grade in this subject.

Exceeding Expectations
A student who performed at this level exceeded grade-level expectations by demonstrating mastery of the subject matter.

The horizontal gray bar shown in the graphics above and below show the range of likely scores your child would receive if he or she took the test multiple times.

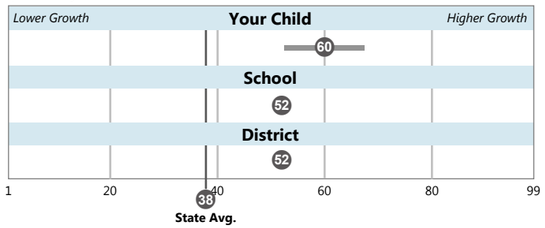
Achievement

How your child performed compared to students in their school, district, and state.

Your Child's Grade	Your Child's Score	Year	Average Score School	Average Score District	Average Score State
5	518	2021	507	507	490
3	526	2019			

2021 Student Growth Percentiles

Your child's score this year is the same as or better than 60 percent of Rhode Island students who had a similar score to your child on the assessment(s) in a previous year(s).



How your child performed in each reporting category and on each individual test question

Reporting Category	Points Earned by Your Child	Total Possible Points	Average Points School	Average Points District	Average Points State	Average Points Earned by Students Meeting Expectations
Reading	22	29	21.7	21.7	17.4	21.5
Language†	10	11	7.0	7.0	5.0	6.1
Writing ‡	5	8	3.2	3.2	2.1	2.5

† The Language reporting category includes the standard English convention scores.
‡ The Writing reporting category is based on the idea development scores.

Individual Test Questions

Question Number	1	2	3	4	5	6	7	8	9	10	11	12 CV	12 ID	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31 CV	31 ID
Points Earned	1/1	1/1	0/2	1/1	1/1	1/1	1/1	1/1	1/1	2/2	2/2	2/3	2/4	1/1	1/1	1/1	1/1	1/1	1/1	0/1	0/1	0/1	1/1	1/1	1/1	1/1	2/2	2/2	0/1	0/1	1/1	3/3	3/4

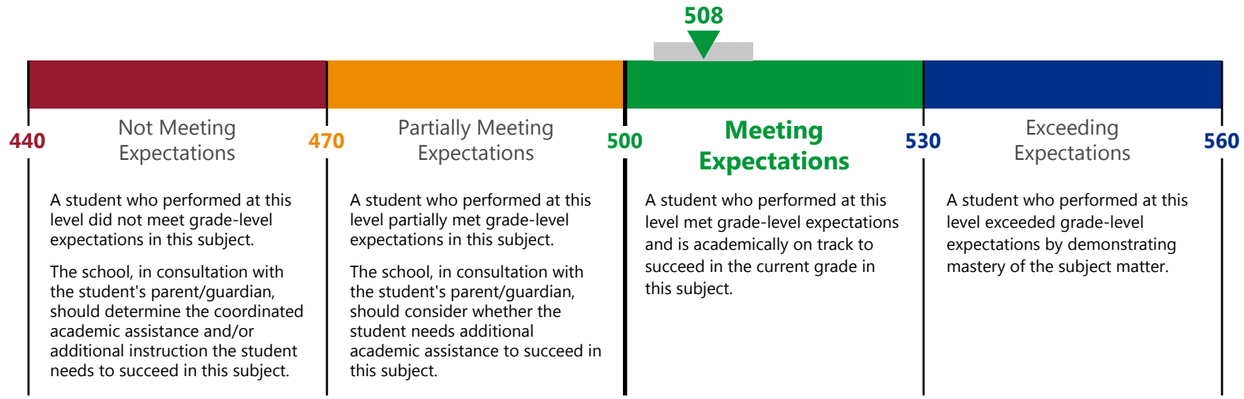
Key x/y = x points earned out of y points possible
Blank space = no answer
N/A = Item not administered
ID = Essay idea development score
CV = Essay conventions score

Name:
SASID:

Grade 5
Spring 2021
Computer-based test

Mathematics

Your Child's Achievement Level:	**Meeting Expectations**
Your Child's Score:	**508**



Not Meeting Expectations
A student who performed at this level did not meet grade-level expectations in this subject.

The school, in consultation with the student's parent/guardian, should determine the coordinated academic assistance and/or additional instruction the student needs to succeed in this subject.

Partially Meeting Expectations
A student who performed at this level partially met grade-level expectations in this subject.

The school, in consultation with the student's parent/guardian, should consider whether the student needs additional academic assistance to succeed in this subject.

Meeting Expectations
A student who performed at this level met grade-level expectations and is academically on track to succeed in the current grade in this subject.

Exceeding Expectations
A student who performed at this level exceeded grade-level expectations by demonstrating mastery of the subject matter.

The horizontal gray bar shown in the graphics above and below show the range of likely scores your child would receive if he or she took the test multiple times.

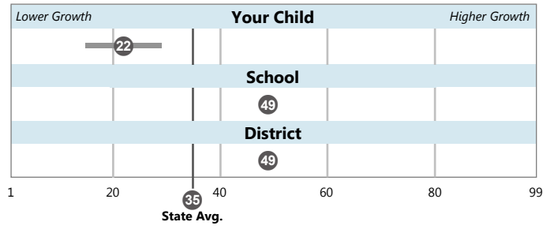
Achievement

How your child performed compared to students in their school, district, and state.

Your Child's Grade	Your Child's Score	Year	Average Score School	Average Score District	Average Score State
5	508	2021	501	501	482
3	528	2019			

2021 Student Growth Percentiles

Your child's score this year is the same as or better than 22 percent of Rhode Island students who had a similar score to your child on the assessment(s) in a previous year(s).



How your child performed in each reporting category and on each individual test question

Reporting Category	Points Earned by Your Child	Total Possible Points	Average Points School	Average Points District	Average Points State	Average Points Earned by Students Meeting Expectations
Operations & Algebraic Thinking	4	8	4.0	4.0	2.8	4.4
Number & Operations in Base Ten	13	16	10.7	10.7	7.0	11.1
Number & Operations-Fractions	10	14	8.1	8.1	4.7	7.5
Measurement & Data	8	10	6.2	6.2	3.8	6.4
Geometry	2	6	2.9	2.9	2.3	3.1

Individual Test Questions

Question Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Points Earned	1/1	1/1	1/1	2/4	0/1	1/1	1/1	0/1	0/1	1/1	1/1	3/4	1/1	1/1	1/1	1/2	1/1	1/1	0/1	1/1	0/1	1/1	1/1	1/1	1/1	3/4	0/1	1/1	1/2	1/1	1/1	1/1	0/1	1/1	4/4	0/1	0/1	0/1	0/1	1/1

Key x/y = x points earned out of y points possible
Blank space = no answer
N/A = Item not administered

APPENDIX F

ITEM-LEVEL CLASSICAL STATISTICS

Table F-1. Item-Level Classical Test Theory Statistics—ELA Grade 3

Item Number	Type	Difficulty	Discrimination	Percent Omitted (%)
EL735534612	MC	0.62	0.47	0
EL735535111	MC	0.47	0.37	0
EL735619614	MC	0.67	0.48	0
EL735629990	MC	0.43	0.42	0
EL735650731	MC	0.77	0.46	0
EL735653115	MC	0.60	0.41	0
EL735655419	MC	0.53	0.35	0
EL735720857	OR	0.75	0.57	0
EL735722199	OR	0.49	0.47	1
EL735726219	OR	0.57	0.53	0
EL735736712#SCORE_TRAIT_Conv	WP	0.25	0.63	3
EL735736712#SCORE_TRAIT_Ideadev	WP	0.10	0.49	3
EL735752249	MC	0.45	0.29	0
EL625955513	MC	0.29	0.23	0
EL625955796	MC	0.58	0.45	0
EL625956196	MC	0.33	0.26	0
EL625956377	MC	0.51	0.37	0
EL625956672	MC	0.59	0.52	0
EL625957401	MC	0.53	0.51	0
EL625959734	MC	0.72	0.39	0
EL625959920	OR	0.62	0.61	0
EL625961096	OR	0.75	0.32	1
EL625962061	OR	0.46	0.29	0
EL625963791	OR	0.34	0.53	2
EL630945115	MC	0.56	0.47	0
EL284943	MC	0.48	0.41	0
EL284944	MC	0.45	0.42	0
EL284946	MC	0.32	0.39	0
EL284948	MC	0.44	0.47	0
EL284949	MC	0.66	0.39	0
EL284950	MC	0.58	0.49	0

Blank values represent no omitted responses on an item, and 0% is a result of rounding for very small values.

Table F-2. Item-Level Classical Test Theory Statistics—ELA Grade 4

Item Number	Type	Difficulty	Discrimination	Percent Omitted (%)
EL812877459	MC	0.78	0.41	0
EL812878059	MC	0.72	0.37	0
EL812878729	MC	0.55	0.47	0
EL812935959	MC	0.71	0.56	0
EL812936582	MC	0.58	0.33	0
EL812937722	MC	0.55	0.40	0
EL812938303	MC	0.52	0.44	0
EL812941713	MC	0.49	0.27	0
EL812943115	MC	0.74	0.38	0
EL812949238#SCORE_TRAIT_Conv	WP	0.33	0.68	2
EL812949238#SCORE_TRAIT_Ideadev	WP	0.21	0.57	2
EL812951483	OR	0.69	0.52	0
EL812952378	OR	0.57	0.53	2
EL809949160	MC	0.72	0.40	0
EL809950008	OR	0.52	0.56	0
EL810000435	MC	0.54	0.32	0
EL810046581	OR	0.34	0.63	1
EL810048797	MC	0.49	0.45	0
EL810055968	OR	0.49	0.54	1
EL810057059	MC	0.57	0.37	0
EL810078292	OR	0.33	0.47	0
EL810079080	MC	0.56	0.42	0
EL810080136	MC	0.48	0.38	0
EL810082669	MC	0.32	0.38	0
EL810084405	MC	0.62	0.45	0
EL810107042	MC	0.56	0.34	0
EL291029	MC	0.58	0.44	0
EL291032	MC	0.62	0.44	0
EL291033	MC	0.73	0.50	0
EL291035	MC	0.51	0.36	0
EL291039	MC	0.61	0.50	0
EL291341	MC	0.60	0.57	0

Blank values represent no omitted responses on an item, and 0% is a result of rounding for very small values.

Table F-3. Item-Level Classical Test Theory Statistics—ELA Grade 5

Item Number	Type	Difficulty	Discrimination	Percent Omitted (%)
EL805937738	MC	0.78	0.50	0
EL805940359	OR	0.60	0.34	0
EL805943442	MC	0.73	0.43	0
EL805945946	MC	0.67	0.42	0
EL805950210	MC	0.77	0.32	0
EL805953548	MC	0.75	0.35	0
EL805955585	OR	0.51	0.43	1
EL805957484	MC	0.54	0.46	0
EL805960800	OR	0.60	0.53	0
EL806031849	MC	0.53	0.36	0
EL806032735	MC	0.57	0.40	0
EL806033603#SCORE_TRAIT_Conv	WP	0.35	0.68	1
EL806033603#SCORE_TRAIT_Ideadev	WP	0.30	0.67	1
EL736467737	MC	0.74	0.27	0
EL736469872	MC	0.69	0.52	0
EL736470482	MC	0.47	0.31	0
EL736471910	MC	0.69	0.38	0
EL736473519	MC	0.39	0.31	0
EL736473790	MC	0.40	0.38	0
EL736474369	MC	0.66	0.24	--
EL736475762	OR	0.59	0.41	0
EL736478536	OR	0.45	0.56	0
EL736478825#SCORE_TRAIT_Conv	WP	0.31	0.68	1
EL736478825#SCORE_TRAIT_Ideadev	WP	0.23	0.69	1
EL827625874	MC	0.61	0.38	0
EL827627427	MC	0.42	0.37	0
EL284550	MC	0.68	0.27	0
EL284551	MC	0.70	0.44	0
EL284552	MC	0.66	0.41	--
EL284554	MC	0.60	0.49	0
EL284557	MC	0.82	0.49	0
EL284560	MC	0.46	0.27	0
EL284561	MC	0.75	0.45	0

Blank values represent no omitted responses on an item, and 0% is a result of rounding for very small values.

Table F-4. Item-Level Classical Test Theory Statistics—ELA Grade 6

Item Number	Type	Difficulty	Discrimination	Percent Omitted (%)
EL806979864	MC	0.70	0.45	0
EL807001596	MC	0.61	0.29	0
EL807002174	MC	0.68	0.34	0
EL807009150	MC	0.70	0.31	0
EL807010236	MC	0.44	0.36	0
EL807011414	MC	0.58	0.35	0
EL807011890	MC	0.65	0.36	0
EL807016586#SCORE_TRAIT_Conv	WP	0.41	0.76	2
EL807016586#SCORE_TRAIT_Ideadev	WP	0.26	0.73	2
EL807061702	MC	0.66	0.23	0
EL807062301	OR	0.43	0.46	0
EL808245411	MC	0.66	0.35	0
EL808246461	OR	0.58	0.30	1
EL735440256#SCORE_TRAIT_Conv	WP	0.39	0.76	2
EL735440256#SCORE_TRAIT_Ideadev	WP	0.24	0.76	2
EL735550535	MC	0.41	0.42	0
EL735554315	MC	0.50	0.49	0
EL735777933	MC	0.61	0.41	0
EL736178377	OR	0.44	0.51	0
EL736179101	MC	0.64	0.46	0
EL736241262	MC	0.45	0.39	0
EL736248371	MC	0.57	0.52	0
EL736249096	MC	0.59	0.40	0
EL736250247	MC	0.49	0.33	0
EL805862435	OR	0.32	0.33	0
EL827430074	MC	0.38	0.13	0
EL302776	MC	0.43	0.39	0
EL302777	MC	0.52	0.43	0
EL302779	MC	0.61	0.41	0
EL302782	MC	0.56	0.52	0
EL302785	MC	0.66	0.38	0
EL302786	MC	0.61	0.46	0
EL302787	MC	0.47	0.33	0
EL302788	MC	0.73	0.49	0

Blank values represent no omitted responses on an item, and 0% is a result of rounding for very small values.

Table F-5. Item-Level Classical Test Theory Statistics—ELA Grade 7

Item Number	Type	Difficulty	Discrimination	Percent Omitted (%)
EL807432481	MC	0.55	0.39	0
EL807433511	MC	0.65	0.45	0
EL807434187	OR	0.45	0.45	0
EL807435581	MC	0.73	0.44	0
EL807437999	MC	0.44	0.38	0
EL807438350	MC	0.45	0.35	0
EL807439180	MC	0.55	0.40	0
EL807443094	MC	0.78	0.48	0
EL807443512	MC	0.76	0.47	0
EL807443849	OR	0.44	0.45	0
EL807445116	MC	0.51	0.28	0
EL807445842	MC	0.52	0.29	0
EL807456720#SCORE_TRAIT_Conv	WP	0.39	0.74	1
EL807456720#SCORE_TRAIT_Ideadev	WP	0.26	0.72	1
EL807349832#SCORE_TRAIT_Conv	WP	0.30	0.73	1
EL807349832#SCORE_TRAIT_Ideadev	WP	0.24	0.69	1
EL807351804	OR	0.42	0.44	0
EL807353731	MC	0.64	0.42	0
EL807354129	MC	0.55	0.42	--
EL807354565	MC	0.60	0.38	--
EL807354764	MC	0.74	0.52	0
EL807355021	MC	0.65	0.47	0
EL807360122	MC	0.56	0.41	0
EL807365314	OR	0.49	0.47	0
EL807365831	MC	0.64	0.36	0
EL807366049	MC	0.64	0.44	0
EL807366496	MC	0.64	0.46	0
EL307948	MC	0.48	0.44	0
EL307962	MC	0.69	0.36	--
EL307963	MC	0.46	0.24	--
EL307970	MC	0.82	0.39	0
EL307971	MC	0.62	0.35	--
EL307973	MC	0.45	0.36	0
EL307974	MC	0.60	0.35	0

Blank values represent no omitted responses on an item, and 0% is a result of rounding for very small values.

Table F-6. Item-Level Classical Test Theory Statistics—ELA Grade 8

Item Number	Type	Difficulty	Discrimination	Percent Omitted (%)
EL809711064	MC	0.60	0.39	0
EL809713456	OR	0.56	0.62	0
EL809734614	MC	0.71	0.46	0
EL809863460	MC	0.68	0.51	0
EL810133273	MC	0.65	0.47	0
EL810222585	OR	0.47	0.47	0
EL810436835	MC	0.53	0.37	0
EL810439521	MC	0.62	0.50	0
EL810456981	MC	0.47	0.33	0
EL810463548#SCORE_TRAIT_Conv	WP	0.42	0.78	2
EL810463548#SCORE_TRAIT_Ideadev	WP	0.28	0.76	2
EL812838757	MC	0.53	0.45	0
EL812849329	MC	0.70	0.47	0
EL810351551	MC	0.58	0.30	0
EL810356239	MC	0.60	0.39	0
EL810357209	OR	0.58	0.57	0
EL810358526	MC	0.60	0.32	0
EL810561207	MC	0.70	0.43	0
EL810561824	MC	0.64	0.50	0
EL810562108	MC	0.63	0.38	0
EL810562694	MC	0.63	0.31	0
EL810563002	MC	0.41	0.30	0
EL810733917#SCORE_TRAIT_Conv	WP	0.36	0.73	2
EL810733917#SCORE_TRAIT_Ideadev	WP	0.24	0.69	2
EL815005831	OR	0.67	0.52	0
EL827744691	MC	0.82	0.51	0
EL302246	MC	0.61	0.38	0
EL302248	MC	0.54	0.52	0
EL302250	MC	0.55	0.46	0
EL302251	MC	0.75	0.45	0
EL302256	MC	0.49	0.34	0
EL302260	MC	0.59	0.38	0
EL303224	MC	0.66	0.34	0
EL303225	MC	0.64	0.42	0

Blank values represent no omitted responses on an item, and 0% is a result of rounding for very small values.

Table F-7. Item-Level Classical Test Theory Statistics—Mathematics Grade 3

Item Number	Type	Difficulty	Discrimination	Percent Omitted (%)
MA299999	MC	0.65	0.56	0
MA252337	MC	0.76	0.47	0
MA261859A	OR	0.48	0.67	1
MA297454	MC	0.56	0.52	0
MA306313	MC	0.51	0.23	0
MA714453A	OR	0.72	0.44	0
MA218578A	OR	0.20	0.41	1
MA293509	OR	0.33	0.56	0
MA303412	MC	0.39	0.17	0
MA311275	MC	0.51	0.42	0
MA311276	MC	0.63	0.41	0
MA735655717	MC	0.63	0.47	0
MA735657470	OR	0.60	0.29	0
MA735663821	OR	0.33	0.50	0
MA735763771	OR	0.40	0.57	1
MA735851787	OR	0.21	0.69	2
MA735954511	OR	0.35	0.58	1
MA736066577	OR	0.37	0.54	0
MA802236949	OR	0.12	0.26	2
MA309747	OR	0.34	0.62	1
MA306297	MC	0.32	0.49	0
MA261818	MC	0.51	0.42	0
MA293494	MC	0.71	0.58	0
MA297399A	OR	0.42	0.75	1
MA306369	MC	0.34	0.29	0
MA310859	MC	0.51	0.55	0
MA310880	MC	0.40	0.26	0
MA285973A	OR	0.70	0.50	0
MA286750A	OR	0.42	0.71	1
MA703072628	OR	0.21	0.34	1
MA713507891	OR	0.61	0.45	0
MA713536927	OR	0.58	0.49	1
MA734752934	OR	0.42	0.47	1
MA735664932	OR	0.42	0.32	2
MA735734045	OR	0.44	0.62	3
MA735765953	OR	0.49	0.56	2
MA735847023	MC	0.37	0.50	0
MA802238054	OR	0.33	0.35	3
MA287674	MC	0.68	0.26	0
MA306285	MC	0.39	0.40	0

Blank values represent no omitted responses on an item, and 0% is a result of rounding for very small values.

Table F-8. Item-Level Classical Test Theory Statistics—Mathematics Grade 4

Item Number	Type	Difficulty	Discrimination	Percent Omitted (%)
MA306940	MC	0.26	0.36	0
MA307079	MC	0.92	0.30	0
MA311568	OR	0.44	0.49	1
MA713631637	OR	0.47	0.67	0
MA229063	OR	0.35	0.56	2
MA303317	OR	0.27	0.52	0
MA704649496	OR	0.61	0.49	0
MA736379417	OR	0.55	0.60	0
MA736381196	OR	0.39	0.60	2
MA800577964	OR	0.65	0.62	0
MA800628900	OR	0.27	0.41	1
MA800629956	OR	0.12	0.27	0
MA800727128	OR	0.45	0.50	0
MA800763292	OR	0.35	0.60	1
MA800780932	OR	0.56	0.72	0
MA801035466	OR	0.39	0.67	1
MA803730594	OR	0.15	0.46	0
MA270627	MC	0.69	0.46	0
MA297973	MC	0.78	0.32	0
MA311558	MC	0.67	0.34	0
MA302483	OR	0.22	0.48	1
MA311554	MC	0.56	0.36	0
MA221898	OR	0.49	0.54	0
MA713673616	OR	0.28	0.52	0
MA714226701	OR	0.48	0.68	1
MA716535935	OR	0.24	0.63	1
MA311574	OR	0.57	0.42	1
MA713629341	OR	0.24	0.45	0
MA736459765	OR	0.20	0.39	0
MA800661015	OR	0.65	0.45	0
MA803742735	MC	0.45	0.43	0
MA803746135	OR	0.53	0.59	0
MA803846674	OR	0.75	0.47	1
MA803956738	OR	0.53	0.58	0
MA247729	MC	0.33	0.53	0
MA287484	OR	0.53	0.67	0
MA299681	OR	0.41	0.49	0
MA303321	MC	0.51	0.61	0
MA306993	MC	0.38	0.39	0
MA311543	MC	0.56	0.37	0

Blank values represent no omitted responses on an item, and 0% is a result of rounding for very small values.

Table F-9. Item-Level Classical Test Theory Statistics—Mathematics Grade 5

Item Number	Type	Difficulty	Discrimination	Percent Omitted (%)
MA297992	MC	0.49	0.35	0
MA298021	MC	0.58	0.58	0
MA624377498	OR	0.37	0.69	1
MA301160	MC	0.28	0.24	0
MA303315	MC	0.62	0.47	0
MA306456	OR	0.73	0.42	0
MA311279	MC	0.56	0.52	0
MA704359650	OR	0.38	0.66	1
MA301157	MC	0.57	0.27	0
MA800652607	OR	0.39	0.47	1
MA800662477	OR	0.21	0.45	0
MA802282875	OR	0.27	0.51	0
MA802285965	OR	0.20	0.55	1
MA802306160	OR	0.53	0.54	1
MA802381243	OR	0.20	0.44	0
MA248869	MC	0.30	0.32	0
MA261200	MC	0.45	0.24	0
MA273791	MC	0.69	0.47	0
MA301605	OR	0.77	0.37	0
MA306435	MC	0.43	0.25	0
MA624376704	OR	0.26	0.63	2
MA262140	MC	0.65	0.43	0
MA624359515	OR	0.51	0.69	1
MA704359410	OR	0.40	0.66	1
MA715102137	OR	0.37	0.58	0
MA715102381	OR	0.48	0.48	0
MA311339A	OR	0.36	0.14	0
MA800650803	MC	0.28	0.28	0
MA801656092	OR	0.36	0.56	2
MA801763240	OR	0.16	0.32	0
MA802284503	OR	0.31	0.54	0
MA803875524	MC	0.49	0.37	0
MA803876799	MC	0.59	0.28	0
MA804579588	OR	0.51	0.50	0
MA808834267	OR	0.20	0.33	1
MA204869	MC	0.55	0.44	0
MA221208	OR	0.40	0.47	1
MA280476	MC	0.60	0.55	0
MA301167	MC	0.18	0.10	0
MA301169	MC	0.29	0.38	0

Blank values represent no omitted responses on an item, and 0% is a result of rounding for very small values.

Table F-10. Item-Level Classical Test Theory Statistics—Mathematics Grade 6

Item Number	Type	Difficulty	Discrimination	Percent Omitted (%)
MA296350	MC	0.19	0.40	0
MA282262	MC	0.45	0.58	0
MA299673	OR	0.45	0.56	0
MA301222	MC	0.37	0.31	0
MA307225	MC	0.86	0.31	0
MA311652	MC	0.34	0.40	0
MA713831396	OR	0.28	0.71	1
MA735778671	OR	0.37	0.54	1
MA736063629	OR	0.28	0.43	2
MA736069855	MC	0.42	0.44	0
MA736364876	OR	0.25	0.36	1
MA736368137	OR	0.52	0.50	0
MA736449649	OR	0.11	0.50	1
MA800166010	MC	0.52	0.31	0
MA800171425	OR	0.34	0.34	0
MA800173241	OR	0.42	0.45	1
MA805179243	OR	0.21	0.21	1
MA805186387	OR	0.13	0.50	0
MA805283567	OR	0.28	0.47	0
MA311694	OR	0.49	0.64	1
MA624254582	OR	0.20	0.37	0
MA703149889	OR	0.55	0.43	0
MA703249688	OR	0.35	0.71	1
MA311660	MC	0.33	0.30	0
MA713679240	OR	0.37	0.61	0
MA713830373	OR	0.49	0.70	1
MA714275582	OR	0.32	0.32	0
MA736071864	MC	0.61	0.39	0
MA736370121	OR	0.27	0.48	1
MA800160765	OR	0.14	0.48	1
MA800162299	MC	0.31	0.35	0
MA800180478	OR	0.37	0.46	0
MA800440516	OR	0.65	0.50	0
MA805100264	MC	0.06	0.20	0
MA805109765	OR	0.27	0.29	1
MA805111931	MC	0.17	0.16	0
MA805280133	OR	0.27	0.25	1
MA272172	MC	0.53	0.48	0
MA307272	MC	0.78	0.18	--
MA311708	OR	0.08	0.48	0

Blank values represent no omitted responses on an item, and 0% is a result of rounding for very small values.

Table F-11. Item-Level Classical Test Theory Statistics—Mathematics Grade 7

Item Number	Type	Difficulty	Discrimination	Percent Omitted (%)
MA272445	MC	0.30	0.54	0
MA259191	MC	0.70	0.49	--
MA306486	MC	0.61	0.51	0
MA703876323	OR	0.15	0.58	0
MA713847883	OR	0.42	0.59	0
MA713848011	OR	0.46	0.55	1
MA282219	MC	0.47	0.44	0
MA306614	MC	0.49	0.24	0
MA311144	OR	0.30	0.73	2
MA713848115	OR	0.14	0.55	0
MA804442802	MC	0.34	0.17	0
MA804636572	OR	0.40	0.55	0
MA804683024	OR	0.37	0.53	0
MA804701799	OR	0.21	0.65	2
MA228065	OR	0.51	0.56	0
MA239637	OR	0.67	0.43	0
MA289832	MC	0.46	0.46	0
MA298205	OR	0.14	0.58	0
MA306615	MC	0.38	0.33	0
MA306636	MC	0.33	0.29	0
MA295745	OR	0.13	0.64	4
MA298180	MC	0.71	0.31	--
MA703881868	OR	0.47	0.65	0
MA713848308	OR	0.13	0.58	0
MA235431	MC	0.21	0.43	0
MA272764	MC	0.54	0.16	0
MA295758	OR	0.34	0.70	2
MA303692	MC	0.33	0.24	0
MA306596	MC	0.37	0.23	0
MA306646	OR	0.33	0.64	0
MA311089	MC	0.26	0.23	0
MA311105	OR	0.18	0.54	0
MA713848086	OR	0.15	0.50	0
MA713848251	OR	0.31	0.32	0
MA713848322	OR	0.13	0.35	0
MA802884644	OR	0.23	0.33	0
MA804635424	OR	0.14	0.48	0
MA272464	MC	0.49	0.47	0
MA290543	OR	0.56	0.50	0
MA303731	MC	0.29	0.48	0

Blank values represent no omitted responses on an item, and 0% is a result of rounding for very small values.

Table F-12. Item-Level Classical Test Theory Statistics—Mathematics Grade 8

Item Number	Type	Difficulty	Discrimination	Percent Omitted (%)
MA252991	MC	0.67	0.17	0
MA311403	OR	0.23	0.54	0
MA283255	MC	0.53	0.43	0
MA704832888	OR	0.18	0.57	1
MA704833231	OR	0.23	0.20	0
MA704855478	OR	0.27	0.69	3
MA715919560	OR	0.13	0.50	1
MA298198	MC	0.27	0.29	0
MA301470	MC	0.47	0.37	0
MA301714	OR	0.23	0.75	4
MA304463	MC	0.26	0.03	0
MA307425	OR	0.15	0.55	1
MA307603	MC	0.58	0.49	--
MA311422	MC	0.48	0.42	0
MA311427	MC	0.52	0.34	0
MA800475061	OR	0.29	0.53	0
MA800475610	OR	0.23	0.49	1
MA804152353	OR	0.47	0.57	0
MA804155665	OR	0.24	0.46	0
MA228379	OR	0.76	0.38	0
MA284198	MC	0.55	0.38	0
MA307585	MC	0.62	0.43	0
MA311437	OR	0.29	0.70	4
MA713930945	OR	0.44	0.64	1
MA715919745	OR	0.51	0.55	0
MA715919853	OR	0.26	0.56	1
MA715919758	OR	0.46	0.29	0
MA800475590	MC	0.73	0.34	0
MA800475640	MC	0.70	0.36	0
MA800675775	OR	0.16	0.35	0
MA800744715	OR	0.14	0.63	2
MA800754030	OR	0.45	0.62	1
MA800974248	OR	0.24	0.57	0
MA804535094	OR	0.25	0.55	0
MA804543815	OR	0.24	0.54	0
MA804576324	MC	0.43	0.40	0
MA229570	MC	0.37	0.29	0
MA287597	MC	0.40	0.24	0
MA297651	OR	0.40	0.50	0
MA307539	MC	0.34	0.31	0

Blank values represent no omitted responses on an item, and 0% is a result of rounding for very small values.

APPENDIX G

SCORE DISTRIBUTIONS

Table G-1. Item-Level Score Distributions for SR and OR Items and WPs—ELA

Grade	Item Number	Total Possible Points	Percent of Students at Score Point					
			0	1	2	3	4	5
3	EL735720857	2	9.38	29.73	60.43	--	--	--
	EL735722199	2	31.20	36.99	30.75	--	--	--
	EL735726219	2	29.83	26.47	43.54	--	--	--
	EL735736712#SCORE_TRAIT_Conv	3	32.95	54.10	7.39	2.26	--	--
	EL735736712#SCORE_TRAIT_Ideadev	4	69.34	17.88	7.46	1.91	0.10	--
	EL625959920	2	31.87	12.81	55.23	--	--	--
	EL625961096	2	3.84	40.93	54.71	--	--	--
	EL625962061	2	32.58	42.49	24.69	--	--	--
	EL625963791	3	22.02	52.97	18.72	3.83	--	--
4	EL812949238#SCORE_TRAIT_Conv	3	29.53	45.05	17.91	5.84	--	--
	EL812949238#SCORE_TRAIT_Ideadev	4	46.77	24.36	20.88	5.46	0.87	
	EL812951483	2	10.00	41.76	47.75	--	--	--
	EL812952378	2	14.75	51.97	31.37	--	--	--
	EL809950008	2	25.73	44.22	30.02	--	--	--
	EL810046581	3	29.39	41.70	22.22	5.54	--	--
	EL810055968	2	24.28	52.42	22.38	--	--	--
	EL810078292	2	58.53	16.23	25.12	--	--	--
5	EL805940359	2	35.99	7.56	56.44	--	--	--
	EL805955585	2	18.38	59.31	21.71	--	--	--
	EL805960800	2	34.47	11.58	53.94	--	--	--
	EL806033603#SCORE_TRAIT_Conv	3	24.64	48.97	19.49	6.06	--	--
	EL806033603#SCORE_TRAIT_Ideadev	4	15.25	56.07	20.61	6.41	0.82	--
	EL736475762	2	31.84	18.12	50.01	--	--	--
	EL736478536	2	39.57	30.35	29.83	--	--	--
	EL736478825#SCORE_TRAIT_Conv	3	31.50	46.72	15.95	4.67	--	--
	EL736478825#SCORE_TRAIT_Ideadev	4	34.30	42.65	16.87	4.07	0.95	--
6	EL807016586#SCORE_TRAIT_Conv	3	26.39	34.82	25.03	12.23	--	--
	EL807016586#SCORE_TRAIT_Ideadev	5	23.67	38.22	22.00	9.36	4.11	1.10
	EL807062301	2	52.91	8.30	38.73	--	--	--
	EL808246461	2	12.30	58.50	28.29	--	--	--
	EL735440256#SCORE_TRAIT_Conv	3	29.45	33.02	22.86	12.56	--	--
	EL735440256#SCORE_TRAIT_Ideadev	5	28.46	33.86	21.84	10.56	2.88	0.28
	EL736178377	2	42.21	27.02	30.66	--	--	--
	EL805862435	2	61.35	12.68	25.94	--	--	--
7	EL807434187	2	49.93	10.47	39.56	--	--	--
	EL807443849	2	50.64	10.04	39.29	--	--	--
	EL807456720#SCORE_TRAIT_Conv	3	30.54	31.81	22.35	13.85	--	--
	EL807456720#SCORE_TRAIT_Ideadev	5	26.86	35.21	22.61	9.16	3.63	1.09
	EL807349832#SCORE_TRAIT_Conv	3	43.12	30.10	16.91	8.39	--	--
	EL807349832#SCORE_TRAIT_Ideadev	5	18.10	53.07	17.76	6.83	2.23	0.52
	EL807351804	2	38.35	40.15	21.47	--	--	--
	EL807365314	2	46.18	8.71	45.09	--	--	--
8	EL809713456	2	29.62	28.55	41.78	--	--	--
	EL810222585	2	24.03	57.18	18.39	--	--	--
	EL810463548#SCORE_TRAIT_Conv	3	26.52	32.02	25.10	14.22	--	--
	EL810463548#SCORE_TRAIT_Ideadev	5	24.59	33.94	22.69	9.55	5.13	1.96
	EL810357209	2	27.77	28.23	43.96	--	--	--
	EL810733917#SCORE_TRAIT_Conv	3	27.06	41.95	19.41	9.14	--	--
	EL810733917#SCORE_TRAIT_Ideadev	5	19.89	49.52	18.35	6.50	2.52	0.79
	EL815005831	2	15.31	35.09	49.54	--	--	--

Table G-2 Item-Level Score Distributions for SR and OR Items—Mathematics

Grade	Item Number	Total Possible Points	Percent of Students at Score Point					
			0	1	2	3	4	5
	MA261859A	3	25.49	28.19	19.85	25.40	--	--
	MA714453A	1	27.36	72.20	--	--	--	--
	MA218578A	1	79.39	20.00	--	--	--	--
	MA293509	1	66.40	33.15	--	--	--	--
	MA735657470	1	39.36	60.47	--	--	--	--
	MA735663821	1	66.27	33.33	--	--	--	--
	MA735763771	1	58.63	39.90	--	--	--	--
	MA735851787	3	53.55	29.74	11.71	3.37	--	--
	MA735954511	1	64.70	34.71	--	--	--	--
	MA736066577	1	62.49	37.05	--	--	--	--
	MA802236949	1	86.35	12.09	--	--	--	--
3	MA309747	1	65.40	33.79	--	--	--	--
	MA297399A	3	25.28	34.34	27.15	12.45	--	--
	MA285973A	1	29.31	70.43			--	--
	MA286750A	3	31.08	25.25	28.57	14.38	--	--
	MA703072628	1	78.36	20.94	--	--	--	--
	MA713507891	1	38.73	61.05	--	--	--	--
	MA713536927	1	40.48	58.09	--	--	--	--
	MA734752934	1	57.41	41.90	--	--	--	--
	MA735664932	1	55.40	42.39	--	--	--	--
	MA735734045	1	53.25	43.91	--	--	--	--
	MA735765953	1	48.24	49.32	--	--	--	--
	MA802238054	1	64.84	32.58	--	--	--	--
	MA311568	1	55.41	44.07	--	--	--	--
	MA713631637	2	33.90	36.60	29.04	--	--	--
	MA229063	1	63.10	35.06	--	--	--	--
	MA303317	1	72.37	27.29	--	--	--	--
	MA704649496	1	38.53	61.39	--	--	--	--
	MA736379417	1	44.49	55.12	--	--	--	--
	MA736381196	1	58.58	39.00	--	--	--	--
	MA800577964	1	35.13	64.60	--	--	--	--
	MA800628900	1	72.37	27.11	--	--	--	--
	MA800629956	1	87.49	12.32	--	--	--	--
4	MA800727128	1	54.95	44.80	--	--	--	--
	MA800763292	1	64.63	34.82	--	--	--	--
	MA800780932	4	15.02	16.06	18.97	26.35	23.15	--
	MA801035466	4	19.68	32.56	24.83	14.41	7.51	--
	MA803730594	1	84.69	15.04	--	--	--	--
	MA302483	1	77.45	22.02	--	--	--	--
	MA221898	1	51.12	48.66	--	--	--	--
	MA713673616	1	71.45	28.41	--	--	--	--
	MA714226701	2	34.90	32.67	31.92	--	--	--
	MA716535935	4	43.85	22.54	25.08	7.02	1.00	--
	MA311574	1	42.36	57.14	--	--	--	--

continued

Grade	Item Number	Total Possible Points	Percent of Students at Score Point					
			0	1	2	3	4	5
4	MA713629341	1	76.09	23.71	--	--	--	--
	MA736459765	1	79.49	20.14	--	--	--	--
	MA800661015	1	34.60	64.91	--	--	--	--
	MA803746135	1	46.41	53.13	--	--	--	--
	MA803846674	1	24.64	74.76	--	--	--	--
	MA803956738	1	46.80	53.03	--	--	--	--
	MA287484	4	8.57	24.26	25.81	26.44	14.53	--
	MA299681	1	58.64	41.26	--	--	--	--
5	MA624377498	4	29.82	22.53	19.77	20.26	6.16	--
	MA306456	1	26.70	73.14	--	--	--	--
	MA704359650	4	39.14	8.87	22.85	13.23	14.68	--
	MA800652607	1	59.68	38.99	--	--	--	--
	MA800662477	1	78.28	21.40	--	--	--	--
	MA802282875	1	72.50	27.38	--	--	--	--
	MA802285965	1	79.49	19.94	--	--	--	--
	MA802306160	2	18.45	55.57	25.44	--	--	--
	MA802381243	1	79.88	19.79	--	--	--	--
	MA301605	1	22.39	77.46	--	--	--	--
	MA624376704	4	39.17	30.52	16.17	10.00	2.38	--
	MA624359515	4	19.25	22.57	14.68	18.82	23.61	--
	MA704359410	2	39.95	39.41	20.05	--	--	--
	MA715102137	1	63.07	36.77	--	--	--	--
	MA715102381	1	52.01	47.54	--	--	--	--
	MA311339A	1	63.82	36.04	--	--	--	--
	MA801656092	1	62.26	36.10	--	--	--	--
	MA801763240	1	84.26	15.54	--	--	--	--
	MA802284503	1	68.98	30.79	--	--	--	--
	MA804579588	1	48.52	51.34	--	--	--	--
	MA808834267	1	79.39	19.79	--	--	--	--
	MA221208	1	59.43	39.99	--	--	--	--
6	MA299673	1	54.39	45.15	--	--	--	--
	MA713831396	4	37.39	31.90	12.82	12.64	4.13	--
	MA735778671	1	62.36	36.84	--	--	--	--
	MA736063629	1	70.03	28.06	--	--	--	--
	MA736364876	1	74.74	24.58	--	--	--	--
	MA736368137	1	47.83	51.71	--	--	--	--
	MA736449649	1	87.70	11.45	--	--	--	--
	MA800171425	1	66.08	33.79	--	--	--	--
	MA800173241	1	57.42	41.83	--	--	--	--
	MA805179243	1	78.26	20.82	--	--	--	--
	MA805186387	2	78.50	16.02	5.11	--	--	--
	MA805283567	1	72.06	27.57	--	--	--	--
	MA311694	4	17.55	22.71	25.50	13.92	19.65	--
	MA624254582	2	67.60	24.47	7.80	--	--	--
	MA703149889	1	45.05	54.51	--	--	--	--
	MA703249688	4	24.49	30.14	26.95	12.26	4.98	--

continued

Grade	Item Number	Total Possible Points	Percent of Students at Score Point					
			0	1	2	3	4	5
6	MA713679240	1	62.29	37.34	--	--	--	--
	MA713830373	4	19.12	20.18	18.86	23.44	16.93	--
	MA714275582	1	67.25	32.30	--	--	--	--
	MA736370121	1	72.13	27.20	--	--	--	--
	MA800160765	1	85.27	13.54	--	--	--	--
	MA800180478	1	62.41	37.47	--	--	--	--
	MA800440516	1	34.69	65.27	--	--	--	--
	MA805109765	1	72.12	26.96	--	--	--	--
	MA805280133	1	71.71	27.33	--	--	--	--
	MA311708	1	91.55	8.19	--	--	--	--
7	MA703876323	2	75.67	17.15	6.78	--	--	--
	MA713847883	1	57.66	42.21	--	--	--	--
	MA713848011	1	53.23	45.92	--	--	--	--
	MA311144	4	32.22	30.47	20.41	11.64	3.19	--
	MA713848115	1	85.61	13.92	--	--	--	--
	MA804636572	1	60.01	39.79	--	--	--	--
	MA804683024	1	63.14	36.79	--	--	--	--
	MA804701799	4	62.61	8.05	10.87	11.07	4.94	--
	MA228065	1	48.41	51.16	--	--	--	--
	MA239637	1	32.60	67.31	--	--	--	--
	MA298205	1	85.24	14.39	--	--	--	--
	MA295745	4	52.92	37.72	2.99	0.68	1.90	--
	MA703881868	2	25.72	54.59	19.63	--	--	--
	MA713848308	1	86.98	12.74	--	--	--	--
	MA295758	4	14.86	51.29	17.26	8.98	6.07	--
	MA306646	1	66.95	32.85	--	--	--	--
	MA311105	1	81.68	17.95	--	--	--	--
	MA713848086	1	84.19	15.48	--	--	--	--
	MA713848251	1	68.30	31.45	--	--	--	--
	MA713848322	1	86.56	13.38	--	--	--	--
	MA802884644	1	76.78	22.97	--	--	--	--
	MA804635424	1	85.92	14.00	--	--	--	--
	MA290543	1	43.87	55.84	--	--	--	--
8	MA311403	1	76.38	23.40	--	--	--	--
	MA704832888	1	81.14	17.64	--	--	--	--
	MA704833231	1	76.63	23.11	--	--	--	--
	MA704855478	4	20.98	58.29	9.65	5.39	3.18	--
	MA715919560	1	86.09	13.13	--	--	--	--
	MA301714	4	52.70	19.35	9.24	7.03	7.83	--
	MA307425	1	83.78	15.37	--	--	--	--
	MA800475061	1	70.40	29.20	--	--	--	--
	MA800475610	1	75.82	23.06	--	--	--	--
	MA804152353	2	35.22	35.22	29.08	--	--	--
	MA804155665	1	75.94	23.76	--	--	--	--
	MA228379	1	23.82	76.01	--	--	--	--
	MA311437	4	39.19	22.03	16.92	11.56	6.24	--

continued

Grade	Item Number	Total Possible Points	Percent of Students at Score Point					
			0	1	2	3	4	5
	MA713930945	4	18.96	20.08	28.28	26.53	5.00	--
	MA715919745	1	48.98	50.53	--	--	--	--
	MA715919853	1	73.12	26.34	--	--	--	--
	MA715919758	1	53.98	45.83	--	--	--	--
	MA800675775	2	73.85	20.25	5.65	--	--	--
8	MA800744715	1	84.12	14.16	--	--	--	--
	MA800754030	1	54.55	44.80	--	--	--	--
	MA800974248	1	75.90	23.97	--	--	--	--
	MA804535094	1	74.97	24.58	--	--	--	--
	MA804543815	1	75.99	23.76	--	--	--	--
	MA297651	1	59.47	40.15	--	--	--	--

APPENDIX H

DIFFERENTIAL ITEM FUNCTIONING RESULTS

Table H-1. Number of Items Classified as "Low" or "High" DIF, Overall and by Group Favored—ELA

Grade	Group		Item Type	Number of Items	Number "Low"			Number "High"		
	Reference	Focal			Total	Favoring		Total	Favoring	
						Reference	Focal		Reference	Focal
3	Male	Female	SR	22	2	1	1	0	0	0
			CR	7	0	0	0	0	0	0
			ES	2	0	0	0	0	0	0
	Not ELL	ELL	SR	22	3	3	0	0	0	0
			CR	7	1	1	0	0	0	0
			ES	2	0	0	0	0	0	0
	Not Economically Disadvantaged	Economically Disadvantaged	SR	22	3	3	0	0	0	0
			CR	7	0	0	0	0	0	0
			ES	2	0	0	0	0	0	0
	White	African American	SR	22	6	5	1	0	0	0
			CR	7	0	0	0	0	0	0
			ES	2	0	0	0	0	0	0
		Hispanic / Latino	SR	22	2	2	0	0	0	0
			CR	7	0	0	0	0	0	0
			ES	2	0	0	0	0	0	0
	Students Without Disabilities	Students with Disabilities	SR	22	0	0	0	0	0	0
			CR	7	0	0	0	0	0	0
			ES	2	0	0	0	0	0	0
	Online	Paper	SR	22	0	0	0	0	0	0
			CR	7	0	0	0	0	0	0
			ES	2	0	0	0	0	0	0
4	Male	Female	SR	24	3	1	2	0	0	0
			CR	6	0	0	0	0	0	0
			ES	2	0	0	0	0	0	0
	Not ELL	ELL	SR	24	3	3	0	0	0	0
			CR	6	1	1	0	0	0	0
			ES	2	0	0	0	0	0	0
	Not Economically Disadvantaged	Economically Disadvantaged	SR	24	0	0	0	0	0	0
			CR	6	0	0	0	0	0	0
			ES	2	0	0	0	0	0	0
	White	African American	SR	24	2	1	1	0	0	0
			CR	6	0	0	0	0	0	0
			ES	2	0	0	0	0	0	0
		Hispanic / Latino	SR	24	1	1	0	0	0	0
			CR	6	0	0	0	0	0	0
			ES	2	0	0	0	0	0	0
	Students Without Disabilities	Students with Disabilities	SR	24	1	1	0	1	1	0
			CR	6	0	0	0	0	0	0
			ES	2	1	1	0	0	0	0
	Online	Paper	SR	24	0	0	0	0	0	0
			CR	6	0	0	0	0	0	0
			ES	2	0	0	0	0	0	0
5	Male	Female	SR	24	2	1	1	0	0	0
			CR	5	0	0	0	0	0	0
			ES	4	1	0	1	0	0	0

continued

Grade	Group		Item Type	Number of Items	Number "Low"			Number "High"		
	Reference	Focal			Total	Favoring		Total	Favoring	
						Reference	Focal		Reference	Focal
5	Not ELL	ELL	SR	24	2	2	0	0	0	0
			CR	5	1	1	0	0	0	0
			ES	4	0	0	0	0	0	0
	Not Economically Disadvantaged	Economically Disadvantaged	SR	24	0	0	0	0	0	0
			CR	5	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0
	White	African American	SR	24	1	1	0	0	0	0
			CR	5	1	1	0	0	0	0
			ES	4	0	0	0	0	0	0
		Hispanic / Latino	SR	24	2	2	0	0	0	0
			CR	5	1	1	0	0	0	0
			ES	4	0	0	0	0	0	0
	Students Without Disabilities	Students with Disabilities	SR	24	0	0	0	0	0	0
			CR	5	0	0	0	0	0	0
			ES	4	1	1	0	0	0	0
	Online	Paper	SR	24	0	0	0	0	0	0
			CR	5	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0
6	Male	Female	SR	26	4	2	2	1	1	0
			CR	4	1	1	0	0	0	0
			ES	4	0	0	0	0	0	0
	Not ELL	ELL	SR	26	4	4	0	1	1	0
			CR	4	1	1	0	0	0	0
			ES	4	0	0	0	0	0	0
	Not Economically Disadvantaged	Economically Disadvantaged	SR	26	0	0	0	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0
	White	African American	SR	26	1	1	0	0	0	0
			CR	4	1	1	0	0	0	0
			ES	4	0	0	0	0	0	0
		Hispanic / Latino	SR	26	2	2	0	0	0	0
			CR	4	1	1	0	0	0	0
			ES	4	0	0	0	0	0	0
	Students Without Disabilities	Students with Disabilities	SR	26	10	9	1	0	0	0
			CR	4	1	1	0	0	0	0
			ES	4	1	1	0	0	0	0
	Online	Paper	SR	26	0	0	0	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0
7	Male	Female	SR	26	4	2	2	1	1	0
			CR	4	1	1	0	0	0	0
			ES	4	0	0	0	0	0	0
	Not ELL	ELL	SR	26	3	2	1	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0
	Not Economically Disadvantaged	Economically Disadvantaged	SR	26	0	0	0	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0

continued

Grade	Group		Item Type	Number of Items	Number "Low"			Number "High"		
	Reference	Focal			Total	Favoring		Total	Favoring	
						Reference	Focal		Reference	Focal
7	White	African American	SR	26	3	3	0	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0
		Hispanic / Latino	SR	26	1	0	1	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0
	Students Without Disabilities	Students with Disabilities	SR	26	7	6	1	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	2	2	0	0	0	0
	Online	Paper	SR	26	0	0	0	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0
8	Male	Female	SR	26	4	2	2	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0
	Not ELL	ELL	SR	26	3	3	0	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0
	Not Economically Disadvantaged	Economically Disadvantaged	SR	26	0	0	0	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0
	White	African American	SR	26	2	2	0	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0
		Hispanic / Latino	SR	26	1	1	0	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0
	Students Without Disabilities	Students with Disabilities	SR	26	4	4	0	0	0	0
			CR	4	1	1	0	0	0	0
			ES	4	3	3	0	0	0	0
	Online	Paper	SR	26	0	0	0	0	0	0
			CR	4	0	0	0	0	0	0
			ES	4	0	0	0	0	0	0

Table H-2. Number of Items Classified as "Low" or "High" DIF, Overall and by Group Favored—Mathematics

Grade	Group		Item Type	Number of Items	Number "Low"			Number "High"		
	Reference	Focal			Total	Favoring		Total	Favoring	
						Reference	Focal		Reference	Focal
3	Male	Female	SR	17	0	0	0	0	0	0
			CR	23	1	1	0	0	0	0
	Not ELL	ELL	SR	17	1	1	0	0	0	0
			CR	23	1	1	0	0	0	0
	Not Economically Disadvantaged	Economically Disadvantaged	SR	17	0	0	0	0	0	0
			CR	23	0	0	0	0	0	0
	White	African American	SR	17	1	1	0	0	0	0
			CR	23	6	5	1	0	0	0
		Hispanic / Latino	SR	17	0	0	0	0	0	0
			CR	23	0	0	0	0	0	0
	Students Without Disabilities	Students with Disabilities	SR	17	1	0	1	0	0	0
			CR	23	0	0	0	0	0	0
	Online	Paper	SR	17	0	0	0	0	0	0
			CR	23	0	0	0	0	0	0
4	Male	Female	SR	11	2	1	1	0	0	0
			CR	29	3	3	0	0	0	0
	Not ELL	ELL	SR	11	1	0	1	0	0	0
			CR	29	2	2	0	0	0	0
	Not Economically Disadvantaged	Economically Disadvantaged	SR	11	0	0	0	0	0	0
			CR	29	0	0	0	0	0	0
	White	African American	SR	11	1	1	0	0	0	0
			CR	29	1	0	1	0	0	0
		Hispanic / Latino	SR	11	1	1	0	0	0	0
			CR	29	0	0	0	0	0	0
	Students Without Disabilities	Students with Disabilities	SR	11	3	0	3	0	0	0
			CR	29	3	3	0	0	0	0
	Online	Paper	SR	11	0	0	0	0	0	0
			CR	29	0	0	0	0	0	0
5	Male	Female	SR	18	3	2	1	0	0	0
			CR	22	2	2	0	0	0	0
	Not ELL	ELL	SR	18	4	2	2	1	1	0
			CR	22	1	1	0	0	0	0
	Not Economically Disadvantaged	Economically Disadvantaged	SR	18	0	0	0	0	0	0
			CR	22	0	0	0	0	0	0
	White	African American	SR	18	3	2	1	0	0	0
			CR	22	1	1	0	0	0	0
		Hispanic / Latino	SR	18	1	1	0	0	0	0
			CR	22	0	0	0	0	0	0
	Students Without Disabilities	Students with Disabilities	SR	18	2	2	0	0	0	0
			CR	22	3	3	0	0	0	0
	Online	Paper	SR	18	0	0	0	0	0	0
			CR	22	0	0	0	0	0	0

continued

Grade	Group		Item Type	Number of Items	Number "Low"			Number "High"		
	Reference	Focal			Total	Favoring		Total	Favoring	
						Reference	Focal		Reference	Focal
6	Male	Female	SR	14	0	0	0	0	0	0
			CR	26	2	2	0	0	0	0
	Not ELL	ELL	SR	14	1	1	0	0	0	0
			CR	26	3	3	0	0	0	0
	Not Economically Disadvantaged	Economically Disadvantaged	SR	14	0	0	0	0	0	0
			CR	26	0	0	0	0	0	0
	White	African American	SR	14	1	1	0	0	0	0
			CR	26	1	0	1	1	1	0
		Hispanic / Latino	SR	14	1	1	0	0	0	0
			CR	26	2	2	0	0	0	0
	Students Without Disabilities	Students with Disabilities	SR	14	2	1	1	0	0	0
			CR	26	2	2	0	1	1	0
	Online	Paper	SR	14	0	0	0	0	0	0
			CR	26	0	0	0	0	0	0
7	Male	Female	SR	17	4	2	2	0	0	0
			CR	23	1	1	0	0	0	0
	Not ELL	ELL	SR	17	1	1	0	0	0	0
			CR	23	2	2	0	0	0	0
	Not Economically Disadvantaged	Economically Disadvantaged	SR	17	0	0	0	0	0	0
			CR	23	0	0	0	0	0	0
	White	African American	SR	17	4	4	0	0	0	0
			CR	23	3	3	0	0	0	0
		Hispanic / Latino	SR	17	1	1	0	0	0	0
			CR	23	0	0	0	0	0	0
	Students Without Disabilities	Students with Disabilities	SR	17	7	5	2	0	0	0
			CR	23	3	3	0	2	2	0
	Online	Paper	SR	17	0	0	0	0	0	0
			CR	23	0	0	0	0	0	0
8	Male	Female	SR	16	2	0	2	0	0	0
			CR	24	2	2	0	1	1	0
	Not ELL	ELL	SR	16	1	1	0	0	0	0
			CR	24	1	1	0	0	0	0
	Not Economically Disadvantaged	Economically Disadvantaged	SR	16	0	0	0	0	0	0
			CR	24	1	1	0	0	0	0
	White	African American	SR	16	2	1	1	0	0	0
			CR	24	2	2	0	0	0	0
		Hispanic / Latino	SR	16	1	1	0	0	0	0
			CR	24	1	1	0	0	0	0
	Students Without Disabilities	Students with Disabilities	SR	16	4	3	1	0	0	0
			CR	24	6	3	3	0	0	0
	Online	Paper	SR	16	0	0	0	0	0	0
			CR	24	0	0	0	0	0	0

APPENDIX I

RELIABILITY

Table I-1: Subgroup Reliabilities—ELA

Grade	Subgroup	Number of Students	Raw Score			Alpha	SEM
			Maximum	Mean	Standard Deviation		
3	All Students	9,068	44	20.87	8.71	0.89	2.84
	ELL	1,347	44	15.14	7.08	0.84	2.82
	Economically Disadvantaged	4,229	44	17.14	7.59	0.86	2.85
	African American	789	44	17.76	7.68	0.86	2.86
	Asian	301	44	23.65	8.45	0.89	2.83
	Hispanic	2,607	44	17.12	7.67	0.86	2.85
	Native American/Alaska Native	61	44	17.36	7.86	0.86	2.98
	White	4,834	44	23.33	8.48	0.89	2.82
	Pacific Islander/Hawaiian	9	--	--	--	--	--
	Multiracial	467	44	20.16	9.02	0.90	2.82
	Male	4,623	44	19.88	8.51	0.89	2.83
	Female	4,429	44	21.93	8.79	0.89	2.85
	Special Education	1,505	44	14.52	7.10	0.85	2.78
4	All Students	9,248	44	22.09	9.28	0.90	2.87
	ELL	1,406	44	15.77	7.85	0.87	2.80
	Economically Disadvantaged	4,249	44	18.15	8.30	0.88	2.84
	African American	897	44	19.03	8.92	0.90	2.83
	Asian	330	44	24.30	8.59	0.89	2.88
	Hispanic	2,474	44	18.17	8.60	0.89	2.84
	Native American/Alaska Native	65	44	18.26	8.26	0.89	2.78
	White	5,002	44	24.51	8.91	0.90	2.87
	Pacific Islander/Hawaiian	17	44	20.59	10.83	0.93	2.90
	Multiracial	449	44	21.72	8.92	0.90	2.89
	Male	4,717	44	21.03	9.16	0.90	2.84
	Female	4,503	44	23.22	9.27	0.90	2.89
	Special Education	1,383	44	13.34	6.71	0.83	2.73
5	All Students	9,448	48	24.52	9.48	0.90	3.04
	ELL	1,579	48	18.07	8.52	0.88	2.98
	Economically Disadvantaged	4,360	48	20.27	8.56	0.88	3.01
	African American	790	48	20.90	8.66	0.88	3.01
	Asian	317	48	27.70	9.14	0.89	2.97
	Hispanic	2,635	48	20.37	8.89	0.89	3.01
	Native American/Alaska Native	69	48	17.22	8.30	0.87	2.97
	White	5,141	48	27.13	8.93	0.89	3.03
	Pacific Islander/Hawaiian	19	48	22.00	8.84	0.87	3.18
	Multiracial	477	48	24.39	9.24	0.89	3.02
	Male	4,817	48	23.62	9.56	0.90	3.02
	Female	4,617	48	25.50	9.28	0.89	3.04
	Special Education	1,471	48	15.02	7.17	0.83	2.92
6	All Students	8,994	50	23.15	10.34	0.90	3.25
	ELL	1,494	50	16.60	8.65	0.87	3.08
	Economically Disadvantaged	3,874	50	18.39	8.84	0.88	3.10
	African American	804	50	18.91	9.20	0.88	3.13
	Asian	314	50	25.16	10.44	0.90	3.27

continued

Grade	Subgroup	Number of Students	Raw Score			Alpha	SEM
			Maximum	Mean	Standard Deviation		
6	Hispanic	2,472	50	18.90	9.12	0.88	3.14
	Native American/Alaska Native	66	50	16.95	8.44	0.88	2.95
	White	4,884	50	26.12	10.10	0.90	3.27
	Pacific Islander/Hawaiian	10	50	19.10	9.34	0.90	3.00
	Multiracial	444	50	21.41	9.93	0.90	3.15
	Male	4,585	50	21.79	10.17	0.90	3.19
	Female	4,384	50	24.64	10.31	0.90	3.27
	Special Education	1,331	50	13.63	6.92	0.82	2.90
7	All Students	9,168	50	23.67	10.44	0.90	3.24
	ELL	1,590	50	16.95	8.61	0.87	3.09
	Economically Disadvantaged	3,998	50	19.02	9.09	0.88	3.12
	African American	834	50	18.90	8.88	0.88	3.10
	Asian	306	50	27.96	9.70	0.89	3.24
	Hispanic	2,587	50	19.19	9.28	0.89	3.13
	Native American/Alaska Native	69	50	17.01	7.79	0.84	3.13
	White	4,915	50	26.73	10.11	0.90	3.26
	Pacific Islander/Hawaiian	12	50	23.75	8.57	0.86	3.26
	Multiracial	438	50	22.69	10.39	0.90	3.24
	Male	4,674	50	22.19	10.18	0.90	3.20
	Female	4,468	50	25.24	10.47	0.90	3.25
	Special Education	1,255	50	13.79	6.79	0.81	2.94
8	All Students	9,109	50	25.24	10.76	0.92	3.10
	ELL	1,623	50	18.74	9.98	0.91	3.00
	Economically Disadvantaged	3,770	50	20.25	9.82	0.91	3.02
	African American	832	50	20.68	9.95	0.91	2.99
	Asian	249	50	29.18	10.83	0.92	3.14
	Hispanic	2,480	50	20.60	9.98	0.91	3.03
	Native American/Alaska Native	67	50	16.96	9.32	0.90	2.93
	White	5,015	50	28.18	10.09	0.91	3.09
	Pacific Islander/Hawaiian	16	50	21.25	10.12	0.90	3.12
	Multiracial	450	50	25.68	11.02	0.92	3.09
	Male	4,693	50	23.28	10.56	0.92	3.07
	Female	4,401	50	27.35	10.57	0.91	3.09
	Special Education	1,294	50	14.83	7.54	0.86	2.86

Table I-2. Subgroup Reliabilities—Mathematics

Grade	Subgroup	Number of Students	Raw Score			Alpha	SEM
			Maximum	Mean	Standard		
3	All Students	9,065	48	19.89	11.25	0.93	3.02
	ELL	1,376	48	13.79	9.09	0.90	2.84
	Economically Disadvantaged	4,217	48	14.81	9.26	0.90	2.88
	African American	789	48	15.26	9.29	0.90	2.89
	Asian	302	48	24.83	11.85	0.93	3.04
	Hispanic	2,619	48	15.11	9.39	0.90	2.90
	Native American/Alaska Native	62	48	15.47	10.84	0.93	2.87
	White	4,817	48	23.15	11.17	0.93	3.05
	Pacific Islander/Hawaiian	10	48	15.00	9.61	0.89	3.17
	Multiracial	466	48	18.30	11.26	0.93	2.95
	Male	4,620	48	19.99	11.45	0.93	3.01
	Female	4,429	48	19.82	11.04	0.92	3.03
	Special Education	1,502	48	12.82	9.23	0.91	2.77
4	All Students	9,253	54	22.41	12.78	0.94	3.26
	ELL	1,416	54	14.64	10.07	0.91	3.01
	Economically Disadvantaged	4,247	54	16.34	10.28	0.91	3.10
	African American	898	54	16.85	10.77	0.92	3.12
	Asian	330	54	27.26	12.90	0.94	3.25
	Hispanic	2,475	54	16.51	10.68	0.92	3.10
	Native American/Alaska Native	65	54	14.65	9.86	0.90	3.07
	White	4,997	54	26.21	12.57	0.93	3.27
	Pacific Islander/Hawaiian	17	54	21.00	13.81	0.95	3.11
	Multiracial	457	54	21.27	12.44	0.93	3.22
	Male	4,735	54	23.00	13.13	0.94	3.26
	Female	4,491	54	21.81	12.37	0.93	3.25
	Special Education	1,382	54	11.97	8.87	0.90	2.84
5	All Students	9,434	54	20.54	11.90	0.91	3.49
	ELL	1,613	54	13.96	9.17	0.88	3.20
	Economically Disadvantaged	4,353	54	15.20	9.41	0.88	3.27
	African American	779	54	15.94	10.16	0.89	3.31
	Asian	321	54	26.50	13.12	0.93	3.52
	Hispanic	2,663	54	15.25	9.48	0.88	3.28
	Native American/Alaska Native	67	54	13.90	9.36	0.88	3.19
	White	5,109	54	23.76	11.90	0.91	3.54
	Pacific Islander/Hawaiian	19	54	17.95	11.51	0.91	3.43
	Multiracial	476	54	20.09	12.33	0.92	3.46
	Male	4,813	54	20.59	12.43	0.92	3.48
	Female	4,608	54	20.51	11.34	0.90	3.50
	Special Education	1,457	54	10.75	7.37	0.85	2.90
6	All Students	8,933	54	18.68	10.91	0.91	3.34
	ELL	1,516	54	12.90	8.58	0.87	3.07
	Economically Disadvantaged	3,838	54	13.71	8.26	0.86	3.11
	African American	798	54	14.09	8.66	0.87	3.13
	Asian	320	54	22.93	13.09	0.93	3.43
	Hispanic	2,446	54	14.15	8.52	0.86	3.15

continued

Grade	Subgroup	Number of Students	Raw Score			Alpha	SEM
			Maximum	Mean	Standard		
	Native American/Alaska Native	62	54	12.66	7.50	0.84	3.03
	White	4,859	54	21.65	11.11	0.91	3.39
	Pacific Islander/Hawaiian	9					
6	Multiracial	439	54	17.20	10.42	0.90	3.29
	Male	4,571	54	18.51	10.95	0.91	3.32
	Female	4,338	54	18.91	10.86	0.90	3.36
	Special Education	1,316	54	10.28	6.67	0.82	2.82
	All Students	9,067	54	16.91	10.91	0.92	3.06
	ELL	1,610	54	11.09	8.10	0.89	2.69
	Economically Disadvantaged	3,936	54	12.16	7.94	0.88	2.76
	African American	833	54	11.65	7.68	0.87	2.73
	Asian	301	54	22.88	12.71	0.93	3.29
	Hispanic	2,571	54	12.38	8.30	0.89	2.78
7	Native American/Alaska Native	68	54	10.24	7.60	0.89	2.50
	White	4,850	54	20.05	11.25	0.92	3.18
	Pacific Islander/Hawaiian	11	54	12.91	8.17	0.87	2.90
	Multiracial	426	54	15.58	10.46	0.92	3.02
	Male	4,628	54	17.51	11.21	0.92	3.07
	Female	4,413	54	16.30	10.55	0.92	3.04
	Special Education	1,239	54	8.60	5.60	0.81	2.45
	All Students	8,985	54	18.74	11.14	0.92	3.24
	ELL	1,622	54	12.89	8.35	0.88	2.91
	Economically Disadvantaged	3,697	54	13.46	7.95	0.86	2.95
	African American	813	54	13.30	8.23	0.87	2.94
	Asian	252	54	24.66	12.98	0.93	3.37
	Hispanic	2,430	54	13.68	8.09	0.87	2.97
8	Native American/Alaska Native	67	54	10.27	4.92	0.70	2.67
	White	4,968	54	21.87	11.36	0.91	3.34
	Pacific Islander/Hawaiian	15	54	15.87	8.43	0.86	3.13
	Multiracial	440	54	19.48	12.15	0.93	3.29
	Male	4,645	54	18.40	11.20	0.92	3.20
	Female	4,326	54	19.14	11.06	0.91	3.28
	Special Education	1,291	54	10.35	5.74	0.78	2.67

Table I-3. Reliabilities by Reporting Categories, Grade, and Content Area—ELA

Grade	Item Reporting Category	Label	Number of Items	Raw Score			Alpha	SEM
				Maximum	Mean	Standard Deviation		
3	1	Reading	21	28	15.42	6.11	0.85	2.34
	2	Language	9	12	5.07	2.66	0.69	1.48
	3	Writing	1	4	0.38	0.71	--	--
4	1	Reading	23	30	16.06	6.59	0.87	2.39
	2	Language	8	10	5.18	2.49	0.71	1.35
	3	Writing	1	4	0.85	0.98	--	--
5	1	Reading	24	29	17.43	6.24	0.84	2.53
	2	Language	7	11	4.99	2.51	0.71	1.35
	3	Writing	2	8	2.10	1.53	0.78	0.71
6	1	Reading	25	29	15.55	6.30	0.84	2.53
	2	Language	7	11	5.10	2.80	0.70	1.54
	3	Writing	2	10	2.51	2.07	0.84	0.83
7	1	Reading	25	29	15.90	6.52	0.84	2.58
	2	Language	7	11	5.32	2.87	0.73	1.50
	3	Writing	2	10	2.45	1.93	0.82	0.82
8	1	Reading	24	28	16.70	6.46	0.87	2.34
	2	Language	8	12	5.99	3.05	0.74	1.55
	3	Writing	2	10	2.56	2.05	0.82	0.88

Table I-4. Reliabilities by Reporting Categories, Grade, and Content Area—Mathematics

Grade	Item Reporting Category	Label	Number of Items	Raw Score			Alpha	SEM
				Maximum	Mean	Standard Deviation		
3	1	Operations and Algebraic Thinking	13	15	6.78	4.15	0.85	1.60
	2	Number and Operations in Base Ten	6	8	2.69	2.08	0.71	1.13
	3	Number and Operations-Fractions	7	9	3.75	2.61	0.74	1.33
	4	Measurement and Data	10	12	5.02	3.10	0.73	1.60
	5	Geometry	4	4	1.65	1.05	0.35	0.85
4	1	Operations and Algebraic Thinking	9	10	3.85	2.54	0.77	1.23
	2	Number and Operations in Base Ten	8	11	5.06	3.08	0.72	1.63
	3	Number and Operations-Fractions	12	16	7.28	4.45	0.85	1.74
	4	Measurement and Data	8	11	4.69	2.79	0.69	1.55
	5	Geometry	3	6	1.54	1.44	0.50	1.01
5	1	Operations and Algebraic Thinking	5	8	2.77	2.18	0.46	1.60
	2	Number and Operations in Base Ten	12	16	7.02	4.18	0.78	1.95
	3	Number and Operations-Fractions	10	14	4.67	3.23	0.76	1.57
	4	Measurement and Data	7	10	3.82	2.68	0.69	1.50
	5	Geometry	6	6	2.26	1.44	0.44	1.08
6	1	Ratios and Proportional Relationships	8	11	4.34	2.56	0.65	1.53
	2	The Number System	8	11	4.78	2.88	0.67	1.65
	3	Expressions and Equations	13	16	6.40	4.00	0.78	1.86
	4	Geometry	4	8	1.44	1.67	0.49	1.19
	5	Statistics and Probability	7	8	1.73	1.62	0.54	1.10
7	1	Ratios and Proportional Relationships	7	11	4.37	2.76	0.76	1.35
	2	The Number System	8	11	3.42	2.56	0.62	1.58
	3	Expressions and Equations	12	13	3.51	2.91	0.77	1.38
	4	Geometry	5	8	2.14	1.76	0.66	1.02
	5	Statistics and Probability	8	11	3.47	2.38	0.63	1.45
8	1	Number System & Expressions/Equations	19	22	7.69	4.53	0.82	1.90
	2	Functions	7	11	3.33	2.58	0.67	1.48
	3	Geometry	12	16	5.36	3.93	0.78	1.83
	4	Statistics and Probability	2	5	2.36	1.43	0.35	1.15