

**California Department of Education
Assessment Development and
Administration Division**



**Standards-based Tests in Spanish
Technical Report
2014–15 Administration**

**Submitted March 23, 2016
Educational Testing Service
Contract No. 5417**

Table of Contents

Acronyms and Initialisms Used in the <i>STS Technical Report</i>	viii
Chapter 1: Introduction	1
Background	1
Test Purpose	1
Test Content	1
Intended Population	1
Intended Use and Purpose of Test Scores	2
Testing Window	3
Significant CAASPP Developments in 2014–15	3
Reduction in Paper Reporting.....	3
Reporting Cluster Data	3
Reduced Program Participation Reporting.....	3
Origin of Demographic Data.....	3
Updated Accessibility Supports.....	3
Individualized Aid Option	4
Limitations of the Assessment	4
Score Interpretation	4
Out-of-Level Testing	4
Score Comparison	4
Groups and Organizations Involved with the CAASPP System	4
State Board of Education	4
California Department of Education	5
Contractor—Educational Testing Service (ETS).....	5
Overview of the Technical Report	5
References	7
Chapter 2: An Overview of STS for RLA Processes	8
Item Development	8
Item Formats.....	8
Item Specifications.....	8
Item Banking.....	8
Item Refresh Rate.....	9
Test Assembly	9
Test Length.....	9
Test Blueprints	9
Content Rules and Item Selection.....	9
Psychometric Criteria.....	10
Test Administration	10
Test Security and Confidentiality.....	10
Procedures to Maintain Standardization	11
Universal Tools, Designated Supports, and Accommodations	12
Non-embedded Supports.....	12
Individualized Aids (now “Unlisted Resources”).....	12
Special Services Summaries	12
Scores	13
Aggregation Procedures	13
Equating	14
Post-Equating	14
Pre-Equating.....	14
Calibration.....	14
Scaling	15
Scoring Table Production.....	15
References	17
Appendix 2.A—STS for RLA Items and Estimated Time Chart	18
Appendix 2.B—Reporting Clusters for Reading/Language Arts	19
Reading/Language Arts Standards Test (Grade Two)	19
Reading/Language Arts Standards Test (Grade Three)	19
Reading/Language Arts Standards Test (Grade Four)	19
Reading/Language Arts Standards Test (Grade Five)	19
Reading/Language Arts Standards Test (Grade Six)	19
Reading/Language Arts Standards Test (Grade Seven).....	20

Reading/Language Arts Standards Test (Grade Eight).....	20
Reading/Language Arts Standards Test (Grade Nine)	20
Reading/Language Arts Standards Test (Grade Ten).....	20
Reading/Language Arts Standards Test (Grade Eleven).....	20
Appendix 2.C—Universal Tools, Designated Supports, and Accommodations for the California Assessment of Student Performance and Progress	21
Appendix 2.D—Special Services Summary Tables	22
Chapter 3: Item Development.....	37
Rules for Item Development.....	37
Item Specifications.....	37
Expected Item Ratio.....	38
Selection of Item Writers	39
Criteria for Selecting Item Writers	39
Item Review Process.....	39
Contractor Review	39
Content Expert Reviews	41
Statewide Pupil Assessment Review Panel	43
Field Testing	43
Stand-alone Field Testing	43
Embedded Field-test Items	44
CDE Data Review.....	44
Item Banking.....	44
References	46
Chapter 4: Test Assembly	47
Test Length.....	47
Rules for Item Selection.....	47
Test Blueprint.....	47
Content Rules and Item Selection.....	47
Psychometric Criteria.....	48
Projected Psychometric Properties of the Assembled Tests.....	49
Rules for Item Sequence and Layout	50
Reference.....	51
Appendix 4.A—Technical Characteristics.....	52
Appendix 4.B—Cluster Targets.....	55
Chapter 5: Test Administration.....	65
Test Security and Confidentiality.....	65
ETS’s Office of Testing Integrity.....	65
Test Development.....	65
Item and Data Review.....	66
Item Banking.....	66
Transfer of Forms and Items to the CDE	66
Security of Electronic Files Using a Firewall	67
Printing and Publishing	67
Test Administration	67
Test Delivery.....	67
Processing and Scoring	68
Data Management	68
Statistical Analysis	69
Reporting and Posting Results.....	69
Student Confidentiality	69
Student Test Results.....	69
Procedures to Maintain Standardization	70
Test Administrators	70
Directions for Administration (DFAs).....	71
CAASPP Paper-Pencil Testing Test Administration Manual	72
Test Operations Management System Manuals	72
Test Booklets	72
Universal Tools, Designated Supports, and Accommodations for Students with Disabilities	72
Identification.....	73
Scoring.....	73
Testing Incidents.....	73
Social Media Security Breaches	73
Testing Improprieties.....	74

References	75
Chapter 6: Performance Standards	76
Background	76
Standard-Setting Procedure	76
Development of Competencies Lists	77
Standard-Setting Methodology	78
Results	79
References	80
Chapter 7: Scoring and Reporting	81
Procedures for Maintaining and Retrieving Individual Scores	81
Scoring and Reporting Specifications	81
Scanning and Scoring	81
Types of Scores and Subscores	82
Raw Score	82
Subscore	82
Scale Score	82
Performance Levels	82
Score Verification Procedures	82
Scoring Key Verification Process	82
Overview of Score Aggregation Procedures	83
Individual Scores	83
Group Scores	85
Reports Produced and Scores for Each Report	85
Types of Score Reports	85
Student Score Report Contents	86
Student Score Report Applications	86
Criteria for Interpreting Test Scores	87
Criteria for Interpreting Score Reports	87
References	88
Appendix 7.A—Scale Score Distribution Tables	89
Appendix 7.B—Demographic Summaries	90
Appendix 7.C—Types of Score Reports	100
Chapter 8: Analyses	101
Samples Used for the Analyses	101
Classical Item Analyses	102
Multiple-Choice Items	102
Reliability Analyses	103
Subgroup Reliabilities and SEMs	104
Conditional Standard Errors of Measurement	104
Decision Classification Analyses	105
Validity Evidence	106
Purpose of the STS for RLA	107
The Constructs to Be Measured	107
Interpretations and Uses of the Scores Generated	108
Intended Test Population(s)	108
Validity Evidence Collected	109
Evidence Based on Response Processes	111
Evidence Based on Internal Structure	111
Evidence Based on Consequences of Testing	113
IRT Analyses	113
Post-equating	113
Pre-Equating	113
Summaries of Scaled IRT <i>b</i> -values	113
Evaluation of Pre-equating	114
Equating Results	114
Differential Item Functioning Analyses	114
References	117
Appendix 8.A—Classical Analyses	119
Appendix 8.B—Reliability Analyses	125
Appendix 8.C—Scaling and Equating Results	137

Chapter 9: Quality Control Procedures	147
Quality Control of Item Development	147
Item Specifications.....	147
Item Writers.....	147
Internal Contractor Reviews.....	147
Assessment Review Panel Review	148
Statewide Pupil Assessment Review Panel Review	148
Data Review of Field-tested Items	148
Quality Control of the Item Bank	149
Quality Control of Test Form Development	149
Quality Control of Test Materials	150
Collecting Test Materials.....	150
Processing Test Materials.....	150
Quality Control of Scanning	150
Quality Control of Image Editing	151
Quality Control of Answer Document Processing and Scoring	151
Accountability of Answer Documents	151
Processing of Answer Documents	151
Scoring and Reporting Specifications	152
Storing Answer Documents.....	152
Quality Control of Psychometric Processes	152
Score Key Verification Procedures	152
Quality Control of Item Analyses and the Equating Process	152
Score Verification Process.....	154
Year-to-Year Comparison Analyses.....	154
Offloads to Test Development.....	154
Quality Control of Reporting	154
Electronic Reporting.....	155
Excluding Student Scores from Summary Reports	155
Reference	156
Chapter 10: Historical Comparisons	157
Base Year Comparisons	157
Examinee Performance	158
Test Characteristics	158
Appendix 10.A—Historical Comparisons Tables, Examinee Performance	160
Appendix 10.B—Historical Comparisons Tables, Test Characteristics	163

Tables

Table 2.1 Scale-Score Ranges for Performance Levels	16
Table 2.C.1 Matrix One Parts 2 and 3: Non-Embedded Supports for the STS for RLA	21
Table 2.D.1 Special Services Summary for RLA, Grades Two and Three	22
Table 2.D.2 Special Services Summary for RLA, Grades Four and Five	25
Table 2.D.3 Special Services Summary for RLA, Grades Six and Seven.....	28
Table 2.D.4 Special Services Summary for RLA, Grades Eight and Nine	31
Table 2.D.5 Special Services Summary for RLA, Grades Ten and Eleven.....	34
Table 4.1 Statistical Targets for STS for RLA Test Assembly	49
Table 4.A.1 Summary of 2015 STS for RLA Projected Technical Characteristics	52
Table 4.A.2 Summary of 2015 STS for RLA Projected Statistical Attributes.....	52
Table 7.1 Mean and Standard Deviation of Raw and Scale Scores for the STS for RLA (Overall Population)	84
Table 7.2 Percentages of Examinees in Each Performance Level	84
Table 7.3 Subgroup Definitions.....	85
Table 7.4 Types of STS for RLA Reports.....	86
Table 7.A.1 Distribution of STS for RLA Scale Scores, Grades Two through Eleven (Overall Population).....	89
Table 7.B.1 Demographic Summary for RLA, Grade Two (Overall Population).....	90
Table 7.B.2 Demographic Summary for RLA, Grade Three (Overall Population)	91
Table 7.B.3 Demographic Summary for RLA, Grade Four (Overall Population)	92
Table 7.B.4 Demographic Summary for RLA, Grade Five (Overall Population).....	93
Table 7.B.5 Demographic Summary for RLA, Grade Six (Overall Population).....	94
Table 7.B.6 Demographic Summary for RLA, Grade Seven (Overall Population)	95
Table 7.B.7 Demographic Summary for RLA, Grade Eight (Overall Population)	96
Table 7.B.8 Demographic Summary for RLA, Grade Nine (Overall Population)	97
Table 7.B.9 Demographic Summary for RLA, Grade Ten (Overall Population)	98
Table 7.B.10 Demographic Summary for RLA, Grade Eleven (Overall Population).....	99

Table 7.C.1 Score Reports Reflecting STS for RLA Results.....	100
Table 8.1 Mean and Median Proportion Correct and Point-Biserial Correlation.....	102
Table 8.2 Reliabilities and SEMs for the STS for RLA	104
Table 8.3 Scale Score CSEM at Performance-level Cut Points	105
Table 8.4 Original Year of Administration for the STS for RLA	111
Table 8.A.1 Item-by-item p -value and Point Biserial for RLA, Grades Two and Three—Current Year and Original Year of Administration.....	119
Table 8.A.2 Item-by-item p -value and Point Biserial for RLA, Grades Four through Six—Current Year and Original Year of Administration.....	120
Table 8.A.3 Item-by-item p -value and Point Biserial for RLA, Grades Seven through Nine—Current Year and Original Year of Administration.....	122
Table 8.A.4 Item-by-item p -value and Point Biserial for RLA, Grades Ten and Eleven—Current Year and Original Year of Administration.....	123
Table 8.B.1 Subscore Reliabilities and Intercorrelations for RLA.....	125
Table 8.B.2 Reliabilities and SEMs by Gender (Male)	127
Table 8.B.3 Reliabilities and SEMs by Gender (Female)	127
Table 8.B.4 Reliabilities and SEMs by Economic Status (Not Disadvantaged).....	127
Table 8.B.5 Reliabilities and SEMs by Economic Status (Disadvantaged)	127
Table 8.B.6 Reliabilities and SEMs by Special Services (No Special Services).....	128
Table 8.B.7 Reliabilities and SEMs by Special Services (Special Services).....	128
Table 8.B.8 Reliabilities and SEMs by Attendance in U.S. Schools (< 12 Months).....	128
Table 8.B.9 Reliabilities and SEMs by Attendance in U.S. Schools (\geq 12 Months).....	128
Table 8.B.10 Subscore Reliabilities and SEM for RLA by Gender/Economic Status	129
Table 8.B.11 Subscore Reliabilities and SEM for RLA by Special Services/Attendance in U.S. Schools	131
Table 8.B.12 Reliability of Classification for RLA, Grade Two.....	133
Table 8.B.13 Reliability of Classification for RLA, Grade Three	133
Table 8.B.14 Reliability of Classification for RLA, Grade Four.....	133
Table 8.B.15 Reliability of Classification for RLA, Grade Five.....	134
Table 8.B.16 Reliability of Classification for RLA, Grade Six	134
Table 8.B.17 Reliability of Classification for RLA, Grade Seven	134
Table 8.B.18 Reliability of Classification for RLA, Grade Eight	135
Table 8.B.19 Reliability of Classification for RLA, Grade Nine	135
Table 8.B.20 Reliability of Classification for RLA, Grade Ten	135
Table 8.B.21 Reliability of Classification for RLA, Grade Eleven	136
Table 8.C.1 Conversion for RLA, Grade Two.....	137
Table 8.C.2 Conversion for RLA, Grade Three	138
Table 8.C.3 Conversion for RLA, Grade Four	139
Table 8.C.4 Conversion for RLA, Grade Five.....	140
Table 8.C.5 Conversion for RLA, Grade Six	141
Table 8.C.6 Conversion for RLA, Grade Seven	142
Table 8.C.7 Conversion for RLA, Grade Eight	143
Table 8.C.8 Conversion for RLA, Grade Nine	144
Table 8.C.9 Conversion for RLA, Grade Ten	145
Table 8.C.10 Conversion for RLA, Grade Eleven	146
Table 10.1 Base Years for the STS for RLA	157
Table 10.A.1 Number of Examinees Tested, Scale Score Means, and Standard Deviations of STS for RLA Across Base Year, 2013, 2014, and 2015.....	160
Table 10.A.2 Percentage of Proficient and Above Across Base Year, 2013, 2014, and 2015	160
Table 10.A.3 Percentage of Advanced Across Base Year, 2013, 2014, and 2015	160
Table 10.A.4 Observed Score Distributions of the STS for RLA Across Base Year, 2013, 2014, and 2015 (Grades Two through Four).....	161
Table 10.A.5 Observed Score Distributions of the STS for RLA Across Base Year, 2013, 2014, and 2015 (Grades Five through Seven).....	161
Table 10.A.6 Observed Score Distributions of the STS for RLA Across Base Year, 2014, and 2015 for RLA (Grades Eight through Eleven).....	162
Table 10.B.1 Average Proportion Correct for Operational Test Items Across Base Year, 2013, 2014, and 2015.....	163
Table 10.B.2 Overall IRT b -values for Operational Test Items Across Base Year, 2013, 2014, and 2015	163
Table 10.B.3 Average Point-Biserial Correlation for Operational Test Items Across Base Year, 2013, 2014, and 2015 ..	164
Table 10.B.4 Score Reliabilities (Cronbach's Alpha) Across Base Year, 2013, 2014, and 2015	164
Table 10.B.5 Standard Error of Measurement (SEM) Across Base Year, 2013, 2014, and 2015.....	164

Figures

Figure 3.1	The ETS Item Development Process for the CAASPP System	37
Figure 4.A.1	Plots of Target Information Function and Projected Information for Total Test and Linking Set for RLA	53
Figure 4.B.1	Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Two	55
Figure 4.B.2	Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Three	56
Figure 4.B.3	Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Four	57
Figure 4.B.4	Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Five	58
Figure 4.B.5	Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Six.....	59
Figure 4.B.6	Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Seven.....	60
Figure 4.B.7	Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Eight.....	61
Figure 4.B.8	Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Nine	62
Figure 4.B.9	Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Ten	63
Figure 4.B.10	Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Eleven.....	64
Figure 6.1	Bookmark Standard-setting Process for the STS	78
Figure 8.1	Decision Accuracy for Achieving a Performance Level.....	106
Figure 8.2	Decision Consistency for Achieving a Performance Level	106

Acronyms and Initialisms Used in the STS Technical Report

ADA	Americans with Disabilities Act	IRT	item response theory
AERA	American Educational Research Association	IT	Information Technology
APA	American Psychological Association	LEA	local educational agency
ARP	Assessment Review Panel	MH DIF	Mantel-Haenszel DIF
CAASPP	California Assessment of Student Performance and Progress	NCME	National Council on Measurement in Education
CAHSEE	California High School Exit Examination	NPS	nonsecure, nonsectarian school
CalTAC	California Technical Assistance Center	OIB	ordered item booklet
CAPA	California Alternate Performance Assessment	OTI	Office of Testing Integrity
CCR	<i>California Code of Regulations</i>	<i>p</i> -value	item proportion correct
CDE	California Department of Education	Pt-Bis	point-biserial correlations
CDS	county/district/school	QC	quality control
CELDT	California English Language Development Test	RLA	reading/language arts
CI	confidence interval	SBE	State Board of Education
CMA	California Modified Assessment	SD	standard deviation
CSEMs	conditional standard errors of measurement	SEM	standard error of measurement
CSTs	California Standards Tests	SFTP	secure file transfer protocol
DFA	<i>Directions for Administration</i>	SGID	School and Grade Identification sheet
DIF	differential item functioning	SKM	score key management
DOK	depth of knowledge	SPAR	Statewide Pupil Assessment Review
EC	<i>Education Code</i>	STAR	Standardized Testing and Reporting
EL	English learner	STS	Standards-based Tests in Spanish
ETS	Educational Testing Service	TIF	test information function
FIA	final item analyses	TOMS	Test Operations Management System
GENASYS	Generalized Analysis System	USDOE	United States Department of Education
ICCs	item characteristic curves	WRMSD	weighted root-mean-square difference
IEP	individualized education program		

Chapter 1: Introduction

Background

In 1997 and 1998, the California State Board of Education (SBE) adopted content standards in four major content areas: English–language arts, mathematics, history–social science, and science. These standards were designed to provide state-level input into instruction curricula and serve as a foundation for the state’s school accountability programs.

In order to measure and evaluate student achievement of the content standards, the state instituted the Standardized Testing and Reporting (STAR) Program. This Program, administered annually as paper-pencil assessments, was authorized in 1997 by state law (Senate Bill 376). In 2013, Assembly Bill 484 was introduced to establish California’s new student assessment system, now known as the California Assessment of Student Performance and Progress (CAASPP). The CAASPP System of assessments replaced the STAR Program. The new assessment system includes computer-based tests for English language arts/literacy and mathematics; and paper-pencil tests in science for the California Standards Tests (CSTs), California Modified Assessment (CMA), and California Alternate Performance Assessment (CAPA), and reading/language arts for the Standards-based Tests in Spanish (STS).

During the 2014–15 administration, the CAASPP System had four components for the paper-pencil tests:

- CSTs for Science, produced for California public schools to assess the California content standards for science in grades five, eight, and ten
- CMA for Science, an assessment of students’ achievement of California’s content standards for science in grades five, eight, and ten, developed for students with an individualized education program (IEP) who meet the CMA eligibility criteria approved by the SBE
- CAPA for Science, produced for students with an IEP and who have significant cognitive disabilities in grades five, eight, and ten and are not able to take the CSTs for Science with accommodations and/or non-embedded accessibility supports or the CMA for Science with accommodations
- STS for Reading/Language Arts (RLA), an optional assessment of students’ achievement of California’s content standards for Spanish-speaking English learners (ELs) that is administered as the CAASPP System’s designated primary language test

Test Purpose

The purpose of the STS program is to permit Spanish-speaking ELs to measure their achievement with respect to California’s content standards in reading/language arts through a primary language test in Spanish. These content standards, approved by the SBE, describe what students should know and be able to do at each grade level.

Test Content

The STS are administered in RLA to students in grades two through eleven.

Intended Population

The STS for RLA are optional, multiple-choice tests that are designed for Spanish-speaking ELs in grades two through eleven. Students in grades five, eight, and ten who take the STS

for RLA are required to also take the CSTs for Science or CMA for Science appropriate to their grade level.

The STS for RLA are taken by Spanish-speaking ELs who have been in school in the United States for less than 12 months or who receive instruction in Spanish. However, all students who are ELs and whose primary language is Spanish are eligible to take the STS for RLA. The two distinct STS populations are the “target” and “nontarget/optional” students.

The target population consists of students receiving instruction in Spanish or students who have attended school in the United States for less than 12 months. These are cumulative, not consecutive, months.

The nontarget/optional population consists of students who receive instruction in English and who have attended school in the United States for 12 cumulative months or longer. In the 2014–15, all STS for RLA test-takers in both the target and optional populations were considered voluntary. (Note: These are *not* the students who took the STS for Non-ELs in Dual-immersion Programs. In 2014–15, the STS for RLA and mathematics could be purchased by local educational agencies (LEAs) and administered to students enrolled in a dual-language immersion program and are either nonlimited English proficient or have been redesignated fluent English proficient. Results were not reported for students who took the STS for Non-ELs in Dual-immersion Programs.) Summary statistics are only reported for all Spanish-speaking ELs who took the STS for RLA in grades two through eleven.

The number of examinees taking the grade-level STS for RLA varies significantly across different grade levels, from approximately 200 in grade eleven to approximately 2,800 for grade two. Approximately 66 to 88 percent of the total test-takers are from the target population, depending on grade level.

Parents may submit a written request to have their child exempted from taking any or all parts of the tests within the CAASPP System. Only students whose parents submit a written request may be exempted from taking the tests (*Education Code [EC] Section 60615*).

Intended Use and Purpose of Test Scores

The results for tests within the CAASPP System are used for two primary purposes, described in sections 60602.5(a) and (a)(4). Sections 60602.5(c) and (d) provide additional background on the tests. (Excerpted from the *EC Section 60602 Web page at <http://www.leginfo.ca.gov/cgi-bin/displaycode?section=edc&group=60001-61000&file=60600-60603>*.)

“60602.5(a) It is the intent of the Legislature in enacting this chapter to provide a system of assessments of pupils that has the primary purposes of assisting teachers, administrators, and pupils and their parents; improving teaching and learning; and promoting high-quality teaching and learning using a variety of assessment approaches and item types. The assessments, where applicable and valid, will produce scores that can be aggregated and disaggregated for the purpose of holding schools and local educational agencies accountable for the achievement of all their pupils in learning the California academic content standards.”

“60602.5(a)(4) Provide information to pupils, parents or guardians, teachers, schools, and local educational agencies on a timely basis so that the information can be used to further the development of the pupil and to improve the educational program.”

“60602.5(c) It is the intent of the Legislature that parents, classroom teachers, other educators, pupil representatives, institutions of higher education, business community members, and the public be involved, in an active and ongoing basis, in the design and implementation of the statewide pupil assessment system and the development of assessment instruments.”

“60602.5(d) It is the intent of the Legislature, insofar as is practically feasible and following the completion of annual testing, that the content, test structure, and test items in the assessments that are part of the statewide pupil assessment system become open and transparent to teachers, parents, and pupils, to assist stakeholders in working together to demonstrate improvement in pupil academic achievement. A planned change in annual test content, format, or design, should be made available to educators and the public well before the beginning of the school year in which the change will be implemented.”

Testing Window

The STS for RLA are administered within a 25-day window which begins 12 instructional days before and ends 12 instructional days after the day on which 85 percent of the instructional year is completed. LEAs may use all or any part of the 25 days for testing but are encouraged to schedule testing over no more than a 10- to 15-day period (California *Code of Regulations [CCR], Title 5, Education, Division 1, Chapter 2, Subchapter 3.75, Article 2, § 855[a][2]*; please note this section of 5 CCR has been updated since the 2014–15 CAASPP administration).

Significant CAASPP Developments in 2014–15

Reduction in Paper Reporting

The Student Score Reports were the only printed reports received after test administration. LEAs were able to download preliminary and final aggregate and individual student data at the LEA and school levels. Student Score Reports were also available as downloadable PDFs.

Reporting Cluster Data

Reporting cluster data were not used in reporting student results to LEAs or test sites, or in Student Score Reports. However, reporting cluster results are included in this technical report.

Reduced Program Participation Reporting

No “EL program participation” category information was collected in the individual student-level data.

Origin of Demographic Data

All student demographic data were derived from the California Longitudinal Pupil Achievement Data System (CALPADS) which caused some demographic fields used for data collection, such as those for student ethnicity/race and primary disability code, to be removed from answer documents. The fields remaining on answer documents are related to student identification and test conditions.

Updated Accessibility Supports

The following non-embedded accessibility supports are no longer available for the STS for RLA:

- The student used an assistive device that did not interfere with the independent work of the student

- The student used a dictionary

Individualized Aid Option

The option to note that an individualized aid was used by the student was included on the answer document.

Limitations of the Assessment

Score Interpretation

Teachers and administrators should not use CAASPP results in isolation to make inferences about instructional needs. In addition, it is important to remember that a single test can provide only limited information. Other relevant information should be considered as well. It is advisable for parents to evaluate their child's strengths and weaknesses in the relevant topics by reviewing local assessments, classroom tests, student grades, classroom work, and teacher recommendations in addition to the child's STS for RLA results (CDE, 2013).

Out-of-Level Testing

Each STS for RLA is designed to measure the content corresponding to a specific grade or course and is appropriate for students in the specific grade or course. Testing below a student's grade is not allowed for the STS for RLA or any test in the CAASPP System; all students are required to take the STS for RLA test for the grade in which they are enrolled. LEAs are advised to review all IEPs to ensure that any provision for testing below a student's grade level has been removed.

Score Comparison

When comparing scale score results for the STS for RLA, the reviewer is limited to comparing results only within the same content area and grade. For example, it is appropriate to compare scores obtained by students and/or schools on the 2014–15 grade three STS for RLA; it would not be appropriate to compare scores obtained on the grade three STS for RLA with those obtained on the grade four STS for RLA. The reviewer may compare results for the same content area and grade, within a school, between schools, or between a school and its district, its county, or the state within the same year or to previous years.

Finally, it is inappropriate to conduct any type of score comparisons (including raw score, percent correct, scale score, or performance level comparisons) between the CSTs and STS from any test administration. Although the STS shares the same test blueprint with the CSTs, they follow an independent procedure for test development and establishment of performance levels; therefore, comparison between STS and CST results is discouraged.

Groups and Organizations Involved with the CAASPP System

State Board of Education

The SBE is the state education agency that sets education policy for kindergarten through grade twelve in the areas of standards, instructional materials, assessment, and accountability. The SBE adopts textbooks for kindergarten through grade eight, adopts regulations to implement legislation, and has the authority to grant waivers of the *EC*.

In addition adopting the rules and regulations for itself, its appointees, and California's public schools, the SBE is also the state educational agency responsible for overseeing California's compliance with programs that meet the requirements of the federal Elementary and Secondary Education Act (and now the Every Student Succeeds Act) and the state's

Public School Accountability Act, which measure the academic performance and growth of schools on a variety of academic metrics. (CDE, 2015)

California Department of Education

The CDE oversees California’s public school system, which is responsible for the education of more than 6,200,000 children and young adults in more than 9,800 schools. California aims to provide a world-class education for all students, from early childhood to adulthood. The Department of Education serves California by innovating and collaborating with educators, schools, parents, and community partners which together, as a team, prepares students to live, work, and thrive in a highly connected world.

Within the CDE, it is the District, School & Innovation Branch that oversees programs promoting innovation and improved student achievement. Programs include oversight of statewide assessments and the collection and reporting of educational data. (CDE, 2016)

Contractor—Educational Testing Service (ETS)

The CDE and the SBE contract with Educational Testing Service (ETS) to develop, administer, and report the CAASPP assessments. ETS has overall responsibility for working with the CDE to implement and maintain an effective assessment system as well as having responsibility for producing and distributing materials, processing the tests, and producing reports. Activities directly conducted by ETS include the following:

- Overall management of the program activities;
- Development of all test items;
- Construction and production of test booklets and related test materials;
- Support and training provided to counties, LEAs, and independently testing charter schools;
- Implementation and maintenance of the Test Operations Management System for orders of materials and pre-identification services;
- Completion of all psychometric activities;
- Production of all scannable test materials;
- Packaging, distribution, and collection of testing materials to LEAs and independently testing charter schools;
- Scanning and scoring of all responses; and
- Production of all score reports and data files of test results.

Overview of the Technical Report

This technical report addresses the characteristics of the STS for RLA administered in spring 2014–15. The technical report contains nine additional chapters as follows:

- Chapter 2 presents a conceptual overview of processes involved in a testing cycle for an STS for RLA form. This includes test construction, test administration, generation of test scores, and dissemination of score reports. Information about the distributions of scores aggregated by subgroups based on demographics and the use of special services is also included, as are the references to various chapters that detail the processes briefly discussed in this chapter.
- Chapter 3 describes the procedures followed during the development of valid STS for RLA items before the 2014–15 administration—in 2014–15, the intact test forms from previous test administrations were reused and there was no new item development. The

chapter also explains the process of field-testing new items and the review of items by contractors and content experts.

- Chapter 4 details the content and psychometric criteria that guided the construction of the STS for RLA forms reused in 2014–15.
- Chapter 5 presents the processes involved in the actual administration of the 2014–15 STS for RLA with an emphasis on efforts made to ensure standardization of the tests. It also includes a detailed section that describes the procedures that were followed by ETS to ensure test security.
- Chapter 6 describes the standard-setting process previously conducted to establish cut scores for newly introduced STS for RLA tests.
- Chapter 7 details the types of scores and score reports that are produced at the end of each administration of the STS for RLA.
- Chapter 8 summarizes the results of the test- and item-level analyses performed during the spring 2014–15 administration of the tests. These include the classical item analyses, the reliability analyses that include assessments of test reliability and the consistency and accuracy of the STS for RLA performance-level classifications, and the procedures designed to ensure the validity of STS for RLA score uses and interpretations. Also discussed in this chapter are item response theory, STS for RLA conversion tables, and the considerations and processes involved in pre-equating.
- Chapter 9 highlights the importance of controlling and maintaining the quality of the STS for RLA.
- Chapter 10 presents historical comparisons of various item- and test-level results for the past three years and the base year for each test, which vary according to grade level.

Each chapter contains summary tables in the body of the text. However, extended appendixes that give more detailed information are provided at the end of the relevant chapters.

References

- California Code of Regulations, Title 5, Education, Division 1, Chapter 2, Subchapter 3.75, Article 2, §§ 853.5 and 855.* Retrieved from <http://www.cde.ca.gov/re/lr/rr/caaspp.asp>
- California Department of Education. (2013). *STAR Program information packet for school district and school staff* (p. 15). Sacramento, CA.
- California Department of Education, EdSource, & the Fiscal Crisis Management Assistance Team. (2014). *Fiscal, demographic, and performance data on California's K–12 schools*. Sacramento, CA: Ed-Data. http://www.ed-data.k12.ca.us/App_Resx/EdDataClassic/fsTwoPanel.aspx?#!bottom=/layouts/EdDataClassic/profile.asp?Tab=1&level=04&reportNumber=16
- California State Board of Education. (2012). SBE agenda for July 2012. <http://www.cde.ca.gov/be/ag/ag/yr12/agenda201207.asp>
- California Department of Education. (2015, May). *State Board of Education Responsibilities*. Retrieved from <http://www.cde.ca.gov/be/ms/po/sberesponsibilities.asp>
- California Department of Education. (2016b, January). *Organization*. Retrieved from <http://www.cde.ca.gov/be/ms/po/sberesponsibilities.asp>

Chapter 2: An Overview of STS for RLA Processes

This chapter provides an overview of the processes involved in a typical test development and administration cycle for the Standards-based Tests in Spanish (STS) for Reading/Language Arts (RLA); these processes are similar to those undertaken to develop the California Standards Tests (CSTs). Also described are the specifications maintained by Educational Testing Service (ETS) to implement each of those processes. In 2014–15, intact forms—i.e., test forms from previous test administrations—from different years were used. Table 8.4 on page 111 lists the forms and their original year of use. All ten tests are considered pre-equated.

The chapter is organized to provide a brief description of each process, followed by a summary of the associated specifications. More details about the specifications and the analyses associated with each process are described in other chapters that are referenced in the sections that follow.

Item Development

Item Formats

All tests of the STS for RLA contain four-option multiple-choice items.

Item Specifications

There was no new item development in 2014–15. Prior to the 2013–14 administration, STS for RLA items were developed to measure California content standards adopted by the state in 1997 and designed to conform to principles of item writing defined by ETS (ETS, 2002). ETS maintained and updated an item specifications document, otherwise known as “item writer guidelines,” for each STS for RLA and used an item utilization plan to guide the development of the items for each content area. Item writing emphasis was determined in consultation with the California Department of Education (CDE).

The item specifications describe the characteristics of the items that should be written to measure each content standard; items of the same type should consistently measure the content standards in the same way. The item specifications helped ensure that the items on the STS for RLA measure the content standards in the same way. To achieve this, the item specifications provided detailed information to item writers who developed items for the STS for RLA.

The items selected for each STS for RLA underwent an extensive item review process that is designed to provide the best standards-based tests possible. Details about the item specifications, the item review process, and the item utilization plan are presented in Chapter 3, starting on page 37.

Item Banking

Before newly developed items were placed in the item bank, ETS prepared them for review by content experts and various external review committees such as the Assessment Review Panels (ARPs), which are described in Chapter 3, starting on page 41; and the Statewide Pupil Assessment Review panel, described in Chapter 3, starting on page 43.

Once the ARP review was complete, the items were placed in the item bank along with the associated information obtained at the review sessions. Items that were accepted by the content experts were updated to a “field-test ready” status. ETS then delivered the items to

the CDE by means of a delivery of the California electronic item bank. Items were subsequently field-tested to obtain information about item performance and item statistics that could be used to assemble operational forms.

The CDE then reviewed those items with their statistical data flagged to determine whether they should be used operationally (see page 44 for more information about the CDE's data review). Any additional updates to item content and statistics were based on data collected from the operational use of the items. However, only the latest content of the item is retained in the bank at any time, along with the administration data from every administration that has included the item.

Further details on item banking are presented on page 44 in Chapter 3.

Item Refresh Rate

Prior to using intact forms in the 2014–15 administration, the item utilization plan required that each year, between 5 and 20 percent of items on an operational form were refreshed; these items remained in the item bank for future use. Because the forms were reused, there were no items refreshed in the 2014–15 administration.

Test Assembly

Test Length

The number of operational items in each STS for RLA varies by content area and grade. There are 65 operational items on the STS for RLA in grades two and three and 75 operational items in grades four through eleven. The considerations used in deciding the test length are described on page 47 in Chapter 4.

Each STS for RLA also includes six field-test items in addition to the operational items. Although there was no new item development for the 2014–15 administration, the field-test items were included as part of the reused forms but did not contribute to students' scores. The total number of items, including operational and field-test items, in each STS for RLA form and the estimated time to complete a test form are presented in Appendix 2.A on page 18.

Test Blueprints

ETS selected all STS for RLA items to conform to the State Board of Education-approved California content standards and test blueprints. The test blueprints for the STS for RLA can be found on the CDE STAR STS Blueprints Web page at <http://www.cde.ca.gov/ta/tg/sr/stsblueprints.asp>.

The test blueprints specify the number of items at the individual standard level. In previous administrations, scores for the STS for RLA items were grouped into subcontent areas referred to as “reporting clusters.” For each STS for RLA reporting cluster, the percentage of questions correctly answered was reported on a student's score report. Although only the total test scale scores are reported and cluster scores are no longer included in the score report in 2014–15, a description of the STS for RLA reporting clusters and the standards that comprise each cluster are provided in Appendix 2.B, which starts on page 19.

Content Rules and Item Selection

Intact forms from different years were reused during the 2014–15 administration. (See Table 8.4 on page 111 for administration years.) In a typical development cycle prior to the 2014–15 administration, test developers followed a number of rules when developing a new test form for a given grade and content area. First and foremost, they selected items that

met the blueprint for that grade and content area. Using an electronic item bank, assessment specialists began by identifying a number of linking items. These were items that had appeared in previous operational test administrations and were then used to equate subsequent (new) test forms. After the linking items were approved, assessment specialists populated the rest of the test form.

Linking items were selected to proportionally represent the full blueprint. Each STS for RLA form was a collection of test items designed for a reliable, fair, and valid measure of student achievement within well-defined course content.

Another consideration was the difficulty of each item. Test developers strived to ensure that there were some easy and some hard items and that there were a number of items in the middle range of difficulty. The detailed rules are presented in Chapter 4, which begins on page 47.

Psychometric Criteria

The staff assessed the projected test characteristics during the preliminary review of the assembled forms. The statistical test targets used to develop the 2014–15 forms and the projected characteristics of the assembled forms are presented starting from page 48 in Chapter 4.

The items in test forms were organized and sequenced differently according to the requirements of the content area. Further details on the arrangement of items during test assembly are also described on page 50 in Chapter 4.

All the forms in the 2014–15 STS for RLA administration were used in prior operational test administrations. See Table 8.4 on page 111 for the list containing the administration in which each STS for RLA was originally administered.

Test Administration

It is of utmost priority to administer the STS for RLA in an appropriate, consistent, secure, confidential, and standardized manner.

Test Security and Confidentiality

All tests within the California Assessment of Student Performance and Progress (CAASPP) System are secure documents. For the STS for RLA administration, every person having access to test materials maintains the security and confidentiality of the tests. ETS's Code of Ethics requires that all test information, including tangible materials (such as test booklets, test questions, test results), confidential files, processes, and activities are kept secure. To ensure security for all tests that ETS develops or handles, ETS maintains an Office of Testing Integrity (OTI). A detailed description of the OTI and its mission is presented in Chapter 5 on page 65.

In the pursuit of enforcing secure practices, ETS and the OTI strive to safeguard the various processes involved in a test development and administration cycle. Those processes are listed below. The practices related to each of the following processes are discussed in detail in Chapter 5, starting on page 65.

- Test development
- Item and data review
- Item banking
- Transfer of forms and items to the CDE

- Security of electronic files using a firewall
- Printing and publishing
- Test administration
- Test delivery
- Processing and scoring
- Data management
- Transfer of scores via secure data exchange
- Statistical analysis
- Reporting and posting results
- Student confidentiality
- Student test results

Procedures to Maintain Standardization

The STS for RLA processes are designed so that the tests are administered and scored in a standardized manner. ETS takes all necessary measures to ensure the standardization of the STS for RLA, as described in this section.

Test Administrators

The STS for RLA are administered in conjunction with the other tests that comprise the CAASPP System. ETS employs personnel who facilitate various processes involved in the standardization of an administration cycle.

Staff at local educational agencies (LEAs) who are central to the processes include LEA CAASPP coordinators, CAASPP test site coordinators, test administrators, proctors, and scribes. The responsibilities of each of the staff members are included in the *CAASPP Paper-Pencil Testing Test Administration Manual* (CDE, 2015a); see page 72 in Chapter 5 for more information.

Test Directions

A series of instructions compiled in detailed manuals is provided to the test administrators. Such documents include, but are not limited to, the following:

Directions for Administration (DFAs)—Manuals used by test administrators to administer the STS for RLA to students to be followed exactly so that all students have an equal opportunity to demonstrate their academic achievement (See page 71 in Chapter 5 for more information.)

CAASPP Paper-Pencil Testing Test Administration Manual—Test administration procedures for LEA CAASPP coordinators and CAASPP test site coordinators (See page 72 in Chapter 5 for more information.)

Test Operations Management System (TOMS) manuals—Instructions for the Web-based modules that allow LEA CAASPP coordinators to set up test administrations, assign tests, and assign test settings; every module has its own user manual with detailed instructions on how to use TOMS (See page 72 in Chapter 5 for more information.)

Universal Tools, Designated Supports, and Accommodations

All public school students participate in the CAASPP System, including students with disabilities and English learners (ELs). Most students with individualized education program (IEPs) take the STS for RLA under standard conditions. However, some students with IEPs may need assistance when taking the STS for RLA. This assistance takes the form of universal tools, designated supports, and accommodations. All students in this category may have test administration directions simplified or clarified.

Appendix 2.C on page 21 presents an adaptation of Matrix One of the “Universal Tools, Designated Supports, and Accommodations for the California Assessment of Student Performance and Progress” (CDE, 2015b). Part 2 of Matrix One, found in Table 2.C.1, includes the non-embedded supports; Part 3, also in Table 2.C.1, includes the non-embedded accessibility supports that can be used for the paper-pencil tests. Appendix 2.C shows only the supports that were allowed for the STS for RLA in 2014–15 and were mapped to STS for RLA answer documents so had data that could be collected. Table 2.C.1 also shows the answer document options in section A3 that are reported in Appendix 2.D and were defined but did not map to a specific universal tool, designated support, or accommodation.

The purpose of universal tools, designated supports, and accommodations in testing is to allow *all* students the opportunity to demonstrate what they know and what they are able to do, rather than give students using them an advantage over other students or artificially inflate their scores. Universal tools, designated supports, and accommodations minimize or remove the barriers that could otherwise prevent students from generating results that reflect their achievement in the content area.

Non-embedded Supports

Non-embedded supports—universal tools, designated supports, and accommodations—do not change the construct being measured. For example, if students used a non-embedded support, such as a large-print version of any CAASPP test, the accommodation does not change what was tested. Accommodations are available to students with documented need; these must be identified, approved, and listed in the student’s IEP or Section 504 plan. The use of non-embedded supports does not change the way scores are reported.

Individualized Aids (now “Unlisted Resources”)

Individualized aids are resources that fundamentally change what is being tested and may interfere with the construct being measured. All individualized aids must be identified, approved, and listed in the student’s IEP or Section 504 plan. Individualized aids, when approved, are marked as option Y in Appendix 2.D. (Note that individualized aids have been subsequently renamed to “unlisted resources.”)

Special Services Summaries

The percentage of students using various universal tools, designated supports, and accommodations during the 2014–15 administration of the STS for RLA is presented in Appendix 2.D, which starts on page 22. The data are organized into three sections within each table. The first section presents the percentages of students using each universal tool, designated supports and/or accommodation in the total testing population. The results for students in special education and not in special education are presented in the second section. The third section presents the results for students enrolled in U.S. schools for less than 12 months and students enrolled in U.S. schools for 12 months or more. The information within each section is presented for the relevant grades.

Scores

The STS for RLA total test raw scores equal the sum of examinees' scores on the operational multiple-choice test items.

Total test raw scores on each STS for RLA are converted to three-digit scale scores using the pre-equating process described starting on page 14. STS for RLA results are reported through the use of these scale scores; the scores range from 150 to 600 for each test. Also reported are performance levels obtained by categorizing the scale score into one of the following levels: far below basic, below basic, basic, proficient, or advanced. Scale scores of 300 and 350 correspond to the cut scores for the basic and proficient performance levels, respectively. The state's target is for all students to score at the proficient or advanced level.

In addition to scale scores for the total content-area test, performance on the associated reporting clusters is also reported. The subscore or reporting cluster score is obtained by summing an examinee's scores on the items in each reporting cluster. That information is reported in terms of a percent-correct score.

Detailed descriptions of STS for RLA scores are found in Chapter 7, which starts on page 81.

Aggregation Procedures

In order to provide meaningful results to the stakeholders, STS for RLA scores for a given grade are aggregated at the school, independently testing charter school, district, county, and state levels. The aggregated scores are generated for both individual students and demographic subgroups. The following sections describe the summary results of individual and demographic subgroup STS for RLA scores aggregated at the state level.

Please note that aggregation is performed on valid scores only, which are cases where examinees met all of the following criteria:

1. Met attemptedness criteria
2. Did not have a parental exemption
3. Did not miss any part of the test due to illness or medical emergency
4. Took the STS for RLA as a designated EL
5. Did not test out of level (grade inappropriate)

Individual Scores

Table 7.1 on page 84 in Chapter 7 offers summary statistics for individual scores aggregated at the state level, describing student performance on each STS for RLA for the overall STS student population. Included in the table are the means and standard deviations of student scores expressed in terms of both raw scores and scale scores; and the raw score means and standard deviations expressed as percentages of the total raw score points in each test. Table 7.2 on page 84 presents the percentages of the overall STS for RLA student population in each performance level.

Demographic Subgroup Scores

Statistics summarizing STS for RLA student performance by content area and grade for selected groups of students are provided in Table 7.B.1 through Table 7.B.10, starting on page 90 in Appendix 7.B, for overall STS for RLA student population. In these tables, students are grouped by demographic characteristics, including gender, country of origin, economic status, length of enrollment in U.S. schools, EL program participation, and need for special education services. The tables show the numbers of students with valid scores in

each group, scale score means and standard deviations, and percent in a performance level, as well as mean percent-correct scores for each reporting cluster for each demographic group. Table 7.3 on page 85 provides definitions for the demographic groups included in the tables.

Equating

Post-Equating

In the years when the new forms were developed prior to the 2014–15 administration, the STS for RLA were equated to a reference form using a linking items nonequivalent groups data collection design and methods based on item response theory (IRT) (Hambleton & Swaminathan, 1985). The “base” or “reference” calibrations for the STS for RLA were established by calibrating samples of item response data from a specific administration, through which item parameter estimates for the items in the reused forms were placed on the reference scale using a set of linking items selected from the previous year. Doing so established a scale to which subsequent item calibrations could be linked.

The procedure used for equating the STS for RLA involved three steps: item calibration, item scaling, and production of scoring tables. Each of those steps, as described below, was applied to all of the grade-level STS for RLA during the tests’ original years of administration. Results were not post-equated for the 2014–15 administration.

Pre-Equating

During the 2014–15 administration, because all the test forms were used in previous operational administrations, pre-equating was conducted prior to administration of the tests. Based on the sample invariant property of IRT, all the item parameter estimates were placed on the reference scale in their previous administrations through the post-equating procedure described above. Item parameters derived in such a manner can be used to create raw-score-to-scale-score conversion tables prior to test administration. Neither calibration nor scaling was implemented in the pre-equating process.

Since all STS for RLA were intact forms without any edits or replacement to items, the conversion tables from previous administrations—when the forms were originally used—are directly applied to the current administration. Table 8.4 on page 111 shows the years the forms were introduced for each test.

Calibration

To obtain item calibrations, a proprietary version of the PARSCALE program was used. The estimation process was constrained by setting a common discrimination value for all items equal to 1.0 / 1.7 (or 0.588) and by setting the lower asymptote for all multiple-choice items to zero. The resulting estimation was equivalent to the Rasch model for multiple-choice items. This approach is in line with the current CST equating and scaling procedures. For the purpose of equating, only the operational items are calibrated for each test.

The PARSCALE calibrations were run in two stages following procedures used with other ETS testing programs. In the first stage, estimation imposed normal constraints on the updated prior-ability distribution. The estimates resulting from this first stage were used as starting values for a second PARSCALE run, in which the subject prior distribution was updated after each expectation maximization (EM) cycle with no constraints. For both stages, the metric of the scale was controlled by the constant discrimination parameters.

Scaling

In the years when the new forms were developed prior to the 2014–15 administration, calibrations of the items were linked to the previously obtained reference scale estimates using linking items and the Stocking and Lord (1983) procedure. In the case of the one-parameter model calibrations, this procedure was equivalent to setting the mean of the new item parameter estimates for the linking set equal to the mean of the previously scaled estimates. As noted earlier, the linking set was a collection of items in a current test form that also appeared in last year’s form and was scaled at that time.

The linking process was carried out iteratively by inspecting differences between the transformed new and old (reference) estimates for the linking items and removing items for which the item difficulty estimates changed significantly. Items with large weighted root-mean-square differences (WRMSDs) between item characteristic curves based on the old and new difficulty estimates were removed from the linking set. The differences were calculated using the following formula:

$$WRMSD = \sqrt{\sum_{j=1}^{n_g} w_j [P_n(\theta_j) - P_r(\theta_j)]^2} \quad (2.1)$$

where,

abilities are grouped into intervals of 0.005 ranging from –3.0 to 3.0,

n_g is the number of intervals/groups,

θ_j is the mean of the ability estimates that fall in interval j ,

w_j is a weight equal to the proportion of estimated abilities from the transformed new form in interval j ,

$P_n(\theta_j)$ is the probability of correct response for the transformed new form item at ability θ_j , and

$P_r(\theta_j)$ is the probability of correct response for the old (reference) form item at ability θ_j .

Based on established procedures, any linking items for which the WRMSD was greater than 0.125 were eliminated from the linking set. This criterion has produced reasonable results over time in similar equating work done with other testing programs at ETS.

Scoring Table Production

Once the new item calibrations for each test were transformed to the base scale after items’ initial administration, IRT procedures were used to transform the new form number-correct scores (raw scores) to their corresponding ability (theta). The ability estimates were then transformed to scale scores through linear transformation.

The procedure is based on the relationship between raw scores and ability (theta). For the STS for RLA, which consist entirely of n multiple-choice items, this is the well-known relationship defined in Lord (1980; equations 4–5):

$$\xi(\theta) = \sum_{i=1}^n P_i(\theta) \quad (2.2)$$

where,

$P_i(\theta)$ is the probability of a correct response to item i at ability θ , and

$\xi(\theta)$ is the corresponding true score.

For each integer score ξ_n on the form after its original use, the procedure was used to first solve for the corresponding ability estimate using equation 2.2. The ability estimates were then expressed in the reporting scale metric by applying linear transformation with the appropriate slope and intercept, using equation 2.3:

$$\text{ScaleScore} = \text{Intercept} + \text{Slope} \times \theta \quad (2.3)$$

where,

θ represents student ability.

The slope and intercept for each STS for RLA were developed from the base forms using equations 2.4 and 2.5 because the basic and proficiency cut scores were required to be equal to 300 and 350, respectively.

$$\text{Slope} = \frac{350 - 300}{\theta_{\text{proficient}} - \theta_{\text{basic}}} \quad (2.4)$$

$$\text{Intercept} = 350 - \theta_{\text{proficient}} \times \left(\frac{350 - 300}{\theta_{\text{proficient}} - \theta_{\text{basic}}} \right) \quad (2.5)$$

where,

$\theta_{\text{proficient}}$ represents theta cut score for proficient on the base scale, and

θ_{basic} represents theta cut score for basic on the base scale.

Complete raw-score-to-scale-score conversion tables for the STS for RLA are presented in Table 8.C.1 through Table 8.C.10 in Appendix 8.C, starting on page 137. The raw scores and corresponding transformed scale scores are also listed in those tables. Data used are from the forms' original administration.

For all of the STS for RLA, regardless of when the form was administered, scale scores were adjusted at both ends of the scale so that the minimum reported scale score was 150 and the maximum reported scale score was 600. Raw scores of zero and perfect raw scores were assigned scale scores of 150 and 600, respectively.

The scale-score ranges defining the various performance levels are presented in Table 2.1.

Table 2.1 Scale-Score Ranges for Performance Levels

STS	Far Below Basic	Below Basic	Basic	Proficient	Advanced
2 RLA	150 – 241	242 – 299	300 – 349	350 – 385	386 – 600
3 RLA	150 – 250	251 – 299	300 – 349	350 – 392	393 – 600
4 RLA	150 – 255	256 – 299	300 – 349	350 – 386	387 – 600
5 RLA	150 – 270	271 – 299	300 – 349	350 – 400	401 – 600
6 RLA	150 – 259	260 – 299	300 – 349	350 – 400	401 – 600
7 RLA	150 – 255	256 – 299	300 – 349	350 – 398	399 – 600
8 RLA	150 – 247	248 – 299	300 – 349	350 – 400	401 – 600
9 RLA	150 – 247	248 – 299	300 – 349	350 – 395	396 – 600
10 RLA	150 – 239	240 – 299	300 – 349	350 – 393	394 – 600
11 RLA	150 – 234	235 – 299	300 – 349	350 – 396	397 – 600

References

- California Department of Education. (2015a). *2015 CAASPP paper-pencil testing test administration manual*. Sacramento, CA. Retrieved from <http://www.caaspp.org/rsc/pdfs/CAASPP.ppt-tam.2015.pdf>
- California Department of Education. (2015b). Universal Tools, Designated Supports, and Accommodations for the California Assessment of Student Performance and Progress. Sacramento, CA. <http://www.cde.ca.gov/ta/tg/ai/caasppmatrix1.asp>
- Educational Testing Service. (2002). *ETS standards for quality and fairness*. Princeton, NJ: Author.
- Hambleton, R. K., and Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston, MA: Kluwer-Nijhoff.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillside, NJ: Lawrence Erlbaum Associates, Inc.
- Stocking, M. L., and Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201–10.

Appendix 2.A—STS for RLA Items and Estimated Time Chart

ITEMS AND ESTIMATED TIME CHART					
Standards-Based Tests in Spanish		Reading Language/Arts			
		Total	Part 1	Part 2	Part 3
Grade 2	Total No. of Items	71			
	Time	150	50	50	50
Grade 3	Total No. of Items	71			
	Time	150	50	50	50
Grade 4	Total No. of Items	81			–
	Time	170	85	85	–
Grade 5	Total No. of Items	81			–
	Time	170	85	85	–
Grade 6	Total No. of Items	81			–
	Time	170	85	85	–
Grade 7	Total No. of Items	81			–
	Time	170	85	85	–
Grade 8	Total No. of Items	81			–
	Time	170	85	85	–
Grade 9	Total No. of Items	81			–
	Time	170	85	85	–
Grade 10	Total No. of Items	81			–
	Time	170	85	85	–
Grade 11	Total No. of Items	81			–
	Time	170	85	85	–

Appendix 2.B—Reporting Clusters for Reading/Language Arts

Reading/Language Arts Standards Test (Grade Two)

Reading

Word Analysis and Vocabulary Development	22 items
Reading Comprehension	15 items
Literary Response and Analysis	6 items

Writing

Written Conventions	14 items
Writing Strategies	8 items

Reading/Language Arts Standards Test (Grade Three)

Reading

Word Analysis and Vocabulary Development	20 items
Reading Comprehension	15 items
Literary Response and Analysis	8 items

Writing

Written Conventions	13 items
Writing Strategies	9 items

Reading/Language Arts Standards Test (Grade Four)

Reading

Word Analysis and Vocabulary Development	18 items
Reading Comprehension	15 items
Literary Response and Analysis	9 items

Writing

Written Conventions	18 items
Writing Strategies	15 items

Reading/Language Arts Standards Test (Grade Five)

Reading

Word Analysis and Vocabulary Development	14 items
Reading Comprehension	16 items
Literary Response and Analysis	12 items

Writing

Written Conventions	17 items
Writing Strategies	16 items

Reading/Language Arts Standards Test (Grade Six)

Reading

Word Analysis and Vocabulary Development	13 items
Reading Comprehension	17 items
Literary Response and Analysis	12 items

Writing

Written Conventions	16 items
Writing Strategies	17 items

Reading/Language Arts Standards Test (Grade Seven)

<i>Reading</i>	
Word Analysis and Vocabulary Development	11 items
Reading Comprehension	18 items
Literary Response and Analysis	13 items
<i>Writing</i>	
Written Conventions	16 items
Writing Strategies	17 items

Reading/Language Arts Standards Test (Grade Eight)

<i>Reading</i>	
Word Analysis and Vocabulary Development	9 items
Reading Comprehension	18 items
Literary Response and Analysis	15 items
<i>Writing</i>	
Written Conventions	16 items
Writing Strategies	17 items

Reading/Language Arts Standards Test (Grade Nine)

<i>Reading</i>	
Word Analysis and Vocabulary Development	8 items
Reading Comprehension	18 items
Literary Response and Analysis	16 items
<i>Writing</i>	
Written Conventions	13 items
Writing Strategies	20 items

Reading/Language Arts Standards Test (Grade Ten)

<i>Reading</i>	
Word Analysis and Vocabulary Development	8 items
Reading Comprehension	18 items
Literary Response and Analysis	16 items
<i>Writing</i>	
Written Conventions	13 items
Writing Strategies	20 items

Reading/Language Arts Standards Test (Grade Eleven)

<i>Reading</i>	
Word Analysis and Vocabulary Development	8 items
Reading Comprehension	19 items
Literary Response and Analysis	17 items
<i>Writing</i>	
Written Conventions	9 items
Writing Strategies	22 items

Appendix 2.C—Universal Tools, Designated Supports, and Accommodations for the California Assessment of Student Performance and Progress

Table 2.C.1 Matrix One Parts 2 and 3: Non-Embedded Supports for the STS for RLA

Option	(U) Universal Tool (D) Designated Support (A) Accommodation	
Answer Document Section A3—Accommodations and Modifications		
B	Pupil marks in paper-pencil test booklet (other than responses including highlighting)	U
C	Scribe	A
F	Leave blank	Unmapped
G	Braille (paper-pencil tests)	A
H	Large-print versions of a paper-pencil test (as available)	A
J	Breaks (Tested over more than one day)	U
K	Breaks (Supervised breaks within a section of the test)	U
L	Administration of the test to the pupil at the most beneficial time of day	D
M	Separate setting	D
N	Leave blank	Unmapped
O	American Sign Language	A
V	Leave blank	Unmapped
X	Leave blank	Unmapped
Y	Individualized aid	–
Z	Read aloud	A
Note: The use of additional accessibility supports can be requested.		

- Universal Tools (U) Are available for all pupils. Pupils may turn the support(s) on/off when embedded as part of the technology platform for the computer-administered CAASPP tests or may choose to use it/them when provided as part of a paper-pencil test.
- Designated Supports (D) Are features that are available for use by any pupil for whom the need has been indicated prior to the assessment, by an educator or group of educators.
- Accommodations (A) For the CAASPP system, eligible pupils shall be permitted to take the tests with accommodations if specified in the pupil’s individualized educational program (IEP) or Section 504 plan.

Appendix 2.D—Special Services Summary Tables

Notes:

1. To improve clarity of tables presented in this section, the columns with total number of students using each service are labeled with the particular grade or test name for which the services were utilized. For example, the column with a heading of “Grade 2” in Table 2.D.1 presents the number of students using various special services on the STS for RLA in grade two. The column with the heading of “Pct. of Total” in the same table represents the percent of students using a service out of the total number of test-takers.
2. The total number of test-takers is the total of students listed under “Any universal tool, desig. support, or accommodation” and those listed under “No universal tool, desig. support, or accommodation.”
3. The sum of the numbers of students across subgroups may not match exactly to the total testing population due to the fact that only valid codes were chosen to identify these subgroups.
4. The notation “N/A” is inserted where frequencies for certain accommodations or supports that are not presented in the data. These accommodations or supports include “B: Marked responses in test booklet,” “J: Breaks (Tested over more than one day),” and “K: Breaks (Had supervised breaks).”

Table 2.D.1 Special Services Summary for RLA, Grades Two and Three

All Students Tested	Grade 2 Number	Grade 2 Pct. of Total	Grade 3	Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	1	0.05%	3	0.18%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	2	0.10%	3	0.18%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	2	0.10%	4	0.24%
<i>Any universal tool, desig. support, or accommodation</i>	3	0.15%	6	0.35%
<i>No Universal tool, desig. support, or accommodation</i>	1,945	99.85%	1,688	99.65%

Students Not in Special Education	Grade 2 Number	Grade 2 Pct. of Total	Grade 3 Number	Grade 3 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	1	0.06%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	1	0.05%	1	0.06%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	1	0.05%	2	0.12%
<i>No</i> Universal tool, desig. support, or accommodation	1,858	99.95%	1,606	99.88%
Students in Special Education	Grade 2 Number	Grade 2 Pct. of Total	Grade 3 Number	Grade 3 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	1	1.12%	2	2.33%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	1	1.12%	2	2.33%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	2	2.25%	4	4.65%
<i>Any</i> Universal tool, desig. support, or accommodation	2	2.25%	4	4.65%
<i>No</i> Universal tool, desig. support, or accommodation	87	97.75%	82	95.35%

Students in U.S. Schools < 12 Months	Grade 2 Number	Grade 2 Pct. of Total	Grade 3 Number	Grade 3 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	1	0.38%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	1	0.38%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	1	0.38%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	260	99.62%	316	100.00%
Students in U.S. Schools ≥ 12 Months	Grade 2 Number	Grade 2 Pct. of Total	Grade 3 Number	Grade 3 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	3	0.22%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	2	0.12%	3	0.22%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	1	0.06%	4	0.29%
<i>Any</i> Universal tool, desig. support, or accommodation	2	0.12%	6	0.44%
<i>No</i> Universal tool, desig. support, or accommodation	1,685	99.88%	1,372	99.56%

Table 2.D.2 Special Services Summary for RLA, Grades Four and Five

All Students Tested	Grade 4 Number	Grade 4 Pct. of Total	Grade 5 Number	Grade 5 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	1	0.14%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	2	0.21%	3	0.41%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	2	0.21%	4	0.54%
<i>Any</i> Universal tool, desig. support, or accommodation	2	0.21%	4	0.54%
<i>No</i> Universal tool, desig. support, or accommodation	950	99.79%	730	99.46%
Students Not in Special Education	Grade 4 Number	Grade 4 Pct. of Total	Grade 5 Number	Grade 5 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	905	100.00%	702	100.00%

Students in Special Education	Grade 4 Number	Grade 4 Pct. of Total	Grade 5 Number	Grade 5 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	1	3.13%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	2	4.26%	3	9.38%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	2	4.26%	4	12.50%
<i>Any</i> Universal tool, desig. support, or accommodation	2	4.26%	4	12.50%
<i>No</i> Universal tool, desig. support, or accommodation	45	95.74%	28	87.50%
Students in U.S. Schools < 12 Months	Grade 4 Number	Grade 4 Pct. of Total	Grade 5 Number	Grade 5 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	324	100.00%	272	100.00%

Students in U.S. Schools ≥ 12 Months	Grade 4 Number	Grade 4 Pct. of Total	Grade 5 Number	Grade 5 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	1	0.22%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	2	0.32%	3	0.65%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	2	0.32%	4	0.87%
<i>Any</i> Universal tool, desig. support, or accommodation	2	0.32%	4	0.87%
<i>No</i> Universal tool, desig. support, or accommodation	626	99.68%	458	99.13%

Table 2.D.3 Special Services Summary for RLA, Grades Six and Seven

All Students Tested	Grade 6 Number	Grade 6 Pct. of Total	Grade 7 Number	Grade 7 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	441	100.00%	463	100.00%
Students Not in Special Education	Grade 6 Number	Grade 6 Pct. of Total	Grade 7 Number	Grade 7 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	436	100.00%	457	100.00%

Students in Special Education	Grade 6 Number	Grade 6 Pct. of Total	Grade 7 Number	Grade 7 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	5	100.00%	6	100.00%
Students in U.S. Schools < 12 Months	Grade 6 Number	Grade 6 Pct. of Total	Grade 7 Number	Grade 7 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	273	100.00%	329	100.00%

Students in U.S. Schools ≥ 12 Months	Grade 6 Number	Grade 6 Pct. of Total	Grade 7 Number	Grade 7 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	168	100.00%	134	100.00%

Table 2.D.4 Special Services Summary for RLA, Grades Eight and Nine

All Students Tested	Grade 8 Number	Grade 8 Pct. of Total	Grade 9 Number	Grade 9 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	462	100.00%	847	100.00%
Students Not in Special Education	Grade 8 Number	Grade 8 Pct. of Total	Grade 9 Number	Grade 9 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	454	100.00%	845	100.00%

Students in Special Education	Grade 8 Number	Grade 8 Pct. of Total	Grade 9 Number	Grade 9 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	8	100.00%	2	100.00%
Students in U.S. Schools < 12 Months	Grade 8 Number	Grade 8 Pct. of Total	Grade 9 Number	Grade 9 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	335	100.00%	731	100.00%

Students in U.S. Schools \geq 12 Months	Grade 8 Number	Grade 8 Pct. of Total	Grade 9 Number	Grade 9 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	127	100.00%	116	100.00%

Table 2.D.5 Special Services Summary for RLA, Grades Ten and Eleven

All Students Tested	Grade 10 Number	Grade 10 Pct. of Total	Grade 11 Number	Grade 11 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	287	100.00%	137	100.00%
Students Not in Special Education	Grade 10 Number	Grade 10 Pct. of Total	Grade 11 Number	Grade 11 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	286	100.00%	135	100.00%

Students in Special Education	Grade 10 Number	Grade 10 Pct. of Total	Grade 11 Number	Grade 11 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	1	100.00%	2	100.00%
Students in U.S. Schools < 12 Months	Grade 10 Number	Grade 10 Pct. of Total	Grade 11 Number	Grade 11 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	194	100.00%	98	100.00%

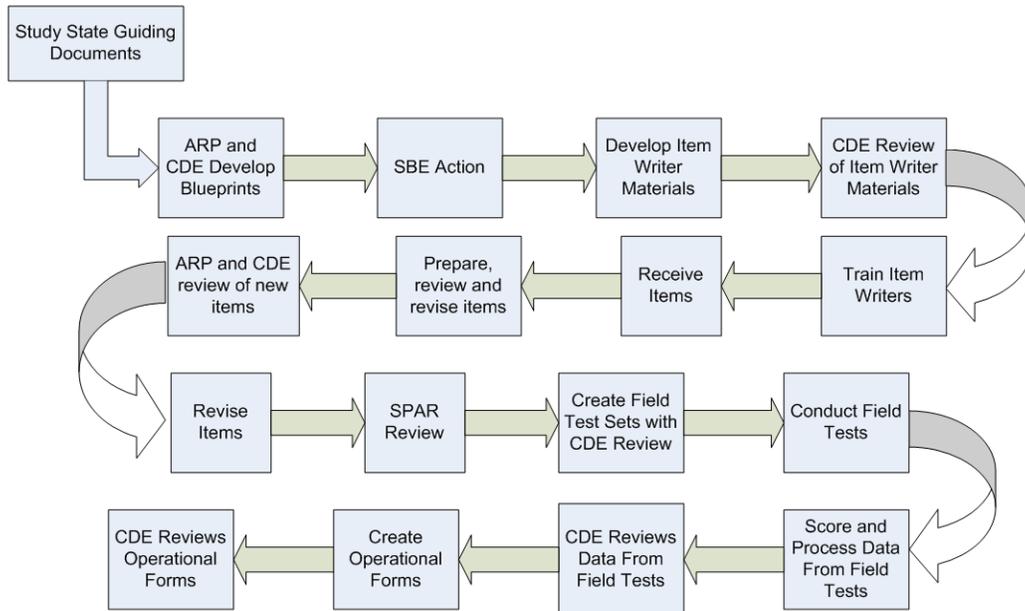
Students in U.S. Schools ≥ 12 Months	Grade 10 Number	Grade 10 Pct. of Total	Grade 11 Number	Grade 11 Pct. of Total
B: Marked in test booklet	N/A	N/A	N/A	N/A
C: Scribe	0	0.00%	0	0.00%
F: Unmapped	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	0	0.00%	0	0.00%
J: Breaks (Tested over more than one day)	N/A	N/A	N/A	N/A
K: Breaks (Had supervised breaks)	N/A	N/A	N/A	N/A
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Unmapped	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Unmapped	0	0.00%	0	0.00%
X: Unmapped	0	0.00%	0	0.00%
Y: Individualized aid	0	0.00%	0	0.00%
Z: Read aloud	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	0	0.00%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	0	0.00%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	93	100.00%	39	100.00%

Chapter 3: Item Development

Intact test forms from previous test administrations from different years were used during the 2014–15 administration. Using an intact form permits the original score conversion tables from the previous administration to be used to look up student scores and performance levels. There was no new item development for the 2014–15 forms.

The Standards-based Tests in Spanish (STS) for Reading/Language Arts (RLA) items were developed to measure California’s content standards and designed to conform to principles of item writing defined by Educational Testing Service (ETS) (ETS, 2002). Each STS for RLA item on the intact forms used in 2014–15 went through a comprehensive development cycle as is described in Figure 3.1 below.

Figure 3.1 The ETS Item Development Process for the CAASPP System



Rules for Item Development

ETS maintained item development specifications for each STS for RLA and developed an item utilization plan to guide the development of the items for each content area. Item writing emphasis was determined in consultation with the California Department of Education (CDE).

Item Specifications

The item specifications described the characteristics of the items that should be written to measure each content standard; items of the same type should consistently measure the content standards in the same way. To achieve this, the item specifications provided detailed information to item writers who developed items for the STS for RLA. The specifications included the following:

- A full statement of each academic content standard, as defined by the State Board of Education (SBE) (CDE, 2009)
- A description of each content strand

- The expected depth of knowledge (DOK) measured by items written for each standard (coded as 1, 2, or 3; items assigned a DOK of 1 are the least cognitively complex and items assigned a DOK of 3 are the most cognitively complex)
- The homogeneity of the construct measured by each standard
- A description of the kinds of item stems appropriate for multiple-choice items used to assess each standard
- A description of the kinds of distractors that are appropriate for multiple-choice items assessing each standard
- A description of appropriate data representations (such as charts, tables, graphs, or other illustrations) for mathematics items
- The content limits for the standard (such as one or two variables, maximum place values of numbers) for mathematics items
- A description of appropriate reading passages, where applicable, for RLA items
- A description of specific kinds of items to be avoided, if any (for example, items with any negative expressions in the stem, e.g. “Which of the following is NOT...”)

In addition, the RLA item specifications contained guidelines for passages used to assess reading comprehension and writing. These guidelines included the following:

- The acceptable ranges for passage length
- The expected distribution of passages by genre
- Guidelines for readability and cognitive load, using standards agreed to by the CDE and ETS
- Expected use of illustrations
- The target number of items that should follow each reading passage and each writing passage
- Appropriate readability levels for reading passages
- A list of topics to be avoided

Expected Item Ratio

ETS prepared the item utilization plan for the development of STS for RLA items. The plan included strategies for developing items that permitted coverage of all appropriate standards for all tests in each content area and at each grade level. ETS test development staff used this plan to determine the number of items to develop for each content area. Because item development has been halted, the item utilization plan is no longer used.

The item utilization plan assumed that the percentage of new items required in an operational form each year matches the development rate listed for each test; these items would remain in the item bank for future use. The plan also declared that an additional five percent of the operational items were likely to be unusable because of normal attrition and noted a need to focus development on “critical” standards, which are those that were difficult to measure well or for which there were few usable items.

It was assumed that at least 75 percent of all field-tested RLA items were expected to have acceptable field-test statistics and become candidates for use in operational tests.

For the 2014–15 STS for RLA administration, field-test items were repeated as a part of the intact forms being reused.

Selection of Item Writers

Criteria for Selecting Item Writers

The items for each STS for RLA were developed by individual item writers with a thorough understanding of the California content standards adopted in 1997. Applicants for item writing were screened by senior ETS content staff. Only those with strong content and teaching backgrounds were approved for inclusion in the training program for item writers. Because most of the participants were current or former California educators, they were particularly knowledgeable about the standards assessed by the STS for RLA. All item writers met the following minimum qualifications:

- Possession of a Bachelor’s degree in the relevant content area or in the field of education with special focus on a particular content of interest; an advanced degree in the relevant content area is desirable
- Previous experience in writing items for standards-based assessments, including knowledge of the many considerations that are important when developing items to match state-specific standards
- Previous experience in writing items in the content areas covered by STS grades and/or courses
- Familiarity, understanding, and support of the California content standards
- Current or previous teaching experience in California, when possible
- Bilingual and biliterate in Spanish and English

Item Review Process

The items selected for each STS for RLA underwent an extensive item review process that was designed to provide the best standards-based tests possible. This section summarizes the various reviews performed that ensure the quality of the STS for RLA items and test forms—currently being reused—at the time the items and forms were developed. See Table 8.4 on page 111 for the dates of the previous administrations.

Contractor Review

Once the items were written, ETS employed a series of internal reviews. The reviews established the criteria used to judge the quality of the item content and were designed to ensure that each item measured what it was intended to measure. The internal reviews also examined the overall quality of the test items before they were prepared for presentation to the CDE and the Assessment Review Panels (ARPs). Because of the complexities involved in producing defensible items for high-stakes programs such as the California Assessment of Student Performance and Progress (CAASPP) System, it was essential that many experienced individuals reviewed each item before it was brought to the CDE, the ARPs, and Statewide Pupil Assessment Review (SPAR) panels.

The ETS review process for the STS for RLA included the following:

1. Internal content review
2. Internal editorial review
3. Internal sensitivity review

Throughout this multistep item review process, the lead content-area assessment specialists and development team members continually evaluated the adherence to the rules for item development.

1. Internal Content Review

Test items and materials underwent two reviews by the content-area assessment specialists. These assessment specialists made sure that the test items and related materials were in compliance with ETS’s written guidelines for clarity, style, accuracy, and appropriateness for California students as well as in compliance with the approved item specifications. Assessment specialists reviewed each item in terms of the following characteristics:

- Relevance of each item to the purpose of the test
- Match of each item to the item specifications, including DOK
- Match of each item to the principles of quality item writing
- Match of each item to the identified standard or standards
- Difficulty of the item
- Accuracy of the content of the item
- Readability of the item or passage
- Grade-level appropriateness of the item
- Appropriateness of any illustrations, graphs, or figures

Each item was classified with a code for the standard it was intended to measure. The assessment specialists checked all items against their classification codes, both to evaluate the correctness of the classification and to ensure that the task posed by the item was relevant to the outcome it was intended to measure. The reviewers could accept the item and classification as written, suggest revisions, or recommend that the item be discarded. These steps occurred prior to the CDE’s review.

2. Internal Editorial Review

After the content-area assessment specialists reviewed each item, a group of specially trained editors also reviewed each item in preparation for consideration by the CDE and the ARPs. The editors checked items for clarity, correctness of language, appropriateness of language for the grade level assessed, adherence to the style guidelines, and conformity with accepted item-writing practices.

3. Internal Sensitivity Review

ETS assessment specialists who are specially trained to identify and eliminate questions that contain content or wording that could be construed to be offensive to or biased against members of specific ethnic, racial, or gender groups, conducted the next level of review. These trained staff members reviewed every item before the CDE and ARP reviews.

The review process promoted a general awareness of and responsiveness to the following:

- Cultural diversity
- Diversity of background, cultural tradition, and viewpoints to be found in the test-taking populations
- Changing roles and attitudes toward various groups
- Role of language in setting and changing attitudes toward various groups
- Contributions of diverse groups (including ethnic and minority groups, individuals with disabilities, and women) to the history and culture of the United States and the achievements of individuals within these groups
- Item accessibility for English-language learners

Content Expert Reviews

Assessment Review Panels

ETS was responsible for working with ARPs as items were developed for the STS for RLA. The ARPs are advisory panels to the CDE and ETS and provided guidance on matters related to item development for the STS for RLA. The ARPs were responsible for reviewing all newly developed items for alignment to the California content standards; these tests use the content standards for English–language arts adopted by the SBE in 1997. The ARPs also reviewed the items for accuracy of content, clarity of phrasing, and quality. In their examination of test items, the ARPs could raise concerns related to age/grade appropriateness and gender, racial, ethnic, and/or socioeconomic bias.

Composition of ARPs

The ARPs comprised current and former teachers, resource specialists, administrators, curricular experts, and other education professionals. Current school staff members met minimum qualifications to serve on the STS ARPs, including:

- Three or more years of general teaching experience in grades kindergarten through twelve and in the content areas (reading/language arts or mathematics);
- Bachelor’s or higher degree in a grade or content area related to reading/language arts or mathematics;
- Knowledge and experience with the California content standards in reading/language arts or mathematics that are current at the time; and
- Bilingual and biliterate in Spanish and English.

School administrators, local educational agency (LEA)/county content/program specialists, or university educators serving on the STS ARPs met the following qualifications:

- Three or more years of experience as a school administrator, LEA/county content/program specialist, or university instructor in a grade-specific area or area related to reading/language arts or mathematics;
- Bachelor’s or higher degree in a grade-specific or subject area related to reading/language arts;
- Knowledge of and experience with the California content standards in reading/language arts that are current at the time; and
- Bilingual and biliterate in Spanish and English.

Every effort was made to ensure that ARP committees included representation of genders and of the geographic regions and ethnic groups in California. Efforts were also made to ensure representation by members with experience working with the diverse student population that makes up STS-eligible test-takers.

ARP members were recruited through an application process. Recommendations were solicited from LEAs and county offices of education as well as from CDE and SBE staff. Applications were reviewed by the ETS assessment directors, who confirmed that the applicant’s qualifications met the specified criteria. Applications that met the criteria are forwarded to CDE and SBE staff for further review and agreement on ARP membership.

ARP Meetings for Review of STS for RLA Items

ETS content-area assessment specialists facilitated the STS for RLA ARP meetings. Each meeting began with a brief training session on how to review items. ETS provided this training, which consisted of the following topics:

- Overview of the purpose and scope of the STS for RLA
- Overview of the STS for RLA's test design specifications and blueprints
- Analysis of the STS for RLA's item specifications
- Overview of criteria for evaluating multiple-choice test items
- Overview of universally accessible Spanish language used to develop multiple-choice test items
- Review and evaluation of items for bias and sensitivity issues

The criteria for evaluating multiple-choice items included the following:

- Overall technical quality
- Match to the California content standards (For the STS for RLA, these are the content standards for science adopted by the SBE in 1997.)
- Match to the construct being assessed by the standard
- Difficulty range
- Clarity
- Correctness of the answer
- Plausibility of the distractors
- Bias and sensitivity factors

Criteria also included more global factors, including—for RLA—the appropriateness, difficulty, and readability of reading passages. The ARPs also were trained on how to make recommendations for revising items.

Guidelines for reviewing items were provided by ETS and approved by the CDE. The set of guidelines for reviewing items is summarized below.

Does the item:

- Have one and only one clearly correct answer?
- Measure the content standard?
- Match the test item specifications?
- Align with the construct being measured?
- Test worthwhile concepts or information?
- Reflect good and current teaching practices?
- Have a stem that gives the student a full sense of what the item is asking?
- Avoid unnecessary wordiness?
- Use response options that relate to the stem in the same way?
- Use response options that are plausible and have reasonable misconceptions and errors?
- Avoid having one response option that is markedly different from the others?
- Avoid clues to students, such as absolutes or words repeated in both the stem and options?
- Reflect content that is free of bias against any person or group?

Is the stimulus, if any, for the item:

- Required in order to answer the item?
- Likely to be interesting to students?
- Clearly and correctly labeled?
- Providing all the information needed to answer the item?

As the first step of the item review process, ARP members reviewed a set of items independently and recorded their individual comments. The next step in the review process was for the group to discuss each item. The content-area assessment specialists facilitated the discussion and recorded all recommendations in a master item review booklet. Item review binders and other item evaluation materials also identified potential bias and sensitivity factors for the ARP to consider as a part of its item reviews.

Depending on CDE approval and the numbers of items still to be reviewed, some ARPs were divided further into smaller groups. These smaller groups were also facilitated by the content-area assessment specialists.

ETS staff maintained the minutes summarizing the review process and then forwarded copies of the minutes to the CDE, emphasizing in particular the recommendations of the panel members.

Statewide Pupil Assessment Review Panel

The SPAR panel is responsible for reviewing and approving all achievement test items to be used statewide for the testing of students in California public schools, grades two through eleven. At the SPAR panel meetings, all new items were presented in binders for review. The SPAR panel representatives ensured that the test items conformed to the requirements of *Education Code* Section 60602. If the SPAR panel rejected specific items, the items were marked for rejection in the item bank and excluded from use on field tests. For the SPAR panel meeting, the item development coordinator was available by telephone to respond to any questions during the course of the meeting.

Field Testing

The primary purposes of field testing are to gather information about item performance and to obtain statistics that can be used to assemble operational forms. However, because intact forms were used with the original field-test items for the 2014–15 CAASPP administration, data were not analyzed for current field-test items.

Stand-alone Field Testing

For each new STS for RLA launched, a pool of items was initially constructed by administering the newly developed items in a stand-alone field test. In stand-alone field testing, examinees were recruited to take tests outside of the usual testing circumstances, and the test results were typically not used for instructional or accountability purposes (Schmeiser & Welch, 2006).

For the STS for RLA in grades eight and above, no stand-alone field testing was conducted due to sample size concerns. Item statistics for these tests were obtained from operational administration and embedded field-testing.

Embedded Field-test Items

Although a stand-alone field test is useful for developing a new test because it can produce a large pool of quality items, embedded field testing is generally preferred because the items being field-tested are seeded throughout the operational test. Variables such as test-taker motivation and test security are the same in embedded field testing as they will be when the field-tested items are later administered operationally.

Such field testing involves distributing the items being field-tested within an operational test form. Different forms contain the same operational items and different field-test items. For the 2014–15 administration, the original field-test items remained in their original positions in the intact forms. Data were not analyzed for field-test items. The numbers of embedded field-test items for the STS for RLA are not presented in this report because for the 2014–15 administration, field-test items were repeated as a part of the intact forms and there was no new item development.

Allocation of Students to Forms

The test forms for a given STS for RLA were spiraled among students in the state so that a large representative sample of test-takers responded to the field-test items embedded in these forms. The spiraling design ensured that a diverse sample of students took each field-test item. The students did not know which items were field-test items and which items were operational items; therefore, their motivation was not expected to vary over the two types of items (Patrick & Way, 2008).

CDE Data Review

Once items were field-tested, ETS prepared the items that failed to meet the desired statistical criteria and the associated statistics for review by the CDE. ETS provided items with their statistical data, along with annotated comment sheets, for the CDE to use in its review. ETS conducted an introductory training to highlight any new issues and serve as a statistical refresher. CDE consultants then made decisions about which items should be included for operational use in the item bank. ETS psychometric and content staff members were available to CDE consultants throughout this process.

Item Banking

Once the ARP new item review is complete, the items were placed in the item bank along with their corresponding review information. Items that were accepted by the ARP, SPAR, and CDE were updated to a “field-test ready” status; items that were rejected were updated to a “rejected before use” status. ETS then delivered the items to the CDE by means of a delivery of the California electronic item bank. Subsequent updates to items were based on field-test and operational use of the items. However, only the latest content of the item is in the bank at any given time, along with the administration data from every administration that included the item.

After field-test or operational use, items that did not meet statistical specifications might be rejected; such items were updated with a status of “rejected for statistical reasons” and remain unavailable in the bank. These statistics were obtained by the psychometrics group at ETS, which carefully evaluated each item for its level of difficulty and discrimination as well as conformance to the item response theory Rasch model. Psychometricians also determined if the item functioned similarly for various subgroups of interest.

All unavailable items were marked with an availability indicator of “Unavailable,” a reason for rejection as described above, and cause alerts so they are not inadvertently included on

subsequent test forms. Statuses and availability were updated programmatically as items were presented for review, accepted or rejected, placed on a form for field-testing, presented for statistical review, and used operationally. All rejection indications were monitored and controlled through ETS's assessment development processes.

ETS currently provides and maintains the electronic item banks for several of the California assessments, including the California High School Exit Examination (CAHSEE), the California English Language Development Test (CELDT), and CAASPP (California Standards Tests for Science, California Modified Assessment for Science, California Alternate Performance Assessment for Science, and STS for RLA). CAHSEE and CAASPP are currently consolidated in the California item banking system. ETS works with the CDE to obtain the data for assessments, such as the CELDT, under contract with other vendors for inclusion into the item bank. ETS provides the item banking application using the local area network architecture and the relational database management system, SQL 2008, already deployed. ETS provides updated versions of the item bank to the CDE on an ongoing basis and works with the CDE to determine the optimum process if a change in databases is desired.

References

- California Department of Education. (2009). *California content standards*. Sacramento, CA. Retrieved from <http://www.cde.ca.gov/be/st/ss/>
- Educational Testing Service. (2002). *ETS standards for quality and fairness*. Princeton, NJ: Author.
- Patrick, R., & Way, D. (March, 2008). *Field testing and equating designs for state educational assessments*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.
- Schmeiser, C. B., & Welch, C. J. (2006). Test development. In R.L. Brennan (Ed.), *Educational measurement* (4th ed.). Westport, CT: American Council on Education and Praeger Publishers.

Chapter 4: Test Assembly

The Standards-based Tests in Spanish (STS) for Reading/Language Arts (RLA) were developed to measure students' performance relative to California's content standards approved by the State Board of Education (SBE) in 1997. They were also constructed to meet professional standards for validity and reliability. For each STS for RLA, the content standards and desired psychometric attributes were used as the basis for assembling the test forms.

Test Length

The number of items in each STS for RLA blueprint was determined by considering the construct that the test is intended to measure and the level of psychometric quality desired. Test length is closely related to the complexity of content to be measured by each test; this content is defined by the California content standards for each grade level and content area. Also considered is the goal that the test be short enough that most of the students complete it in a reasonable amount of time.

The number of operational items on each STS for RLA varies across grades. There are 65 operational items on the STS for RLA in grades two and three and 75 operational items on the STS for RLA in grades four through eleven.

The total number of items also varies. There are a total of 71 items on the STS for RLA in grades two and three and a total of 81 items on the STS for RLA in grades four through eleven.

In addition to operational items, a certain number of the items on each test are field-test items. Among the items on each STS for RLA, six items are field-test items. For more details on the distribution of items, see Appendix 2.A—STS for RLA Items and Time Chart on page 18.

Rules for Item Selection

Test Blueprint

All test items on STS for RLA forms were selected to conform to the SBE-approved California content standards and test blueprints. The content blueprints for the STS for RLA can be found on the California Department of Education (CDE) STAR STS Blueprints Web page at <http://www.cde.ca.gov/ta/tg/sr/stsblueprints.asp>.

Although the test blueprints called for the number of items at the individual standard level, scores for the STS for RLA items are grouped into subcontent areas (reporting clusters). For each STS for RLA reporting cluster, the percentage of questions correctly answered is reported on a student's score report. A list of the STS for RLA reporting clusters by test and the number of items in the cluster that appear in each test are provided in Appendix 2.B—Reporting Clusters, which starts on page 19.

Content Rules and Item Selection

Intact test forms from different years of administration were used during the 2014–15 administration. Prior to the 2013–14 administration, test developers followed a number of rules when developing a new test form for a given grade and content area. First and foremost, they selected items that met the blueprint for that grade level and content area. Using an electronic item bank, assessment specialists began by identifying a number of linking items. These are items that appeared in a previous year's operational administration

and were used to equate the administered test forms. Linking items were selected to proportionally represent the full blueprint. For example, if 25 percent of all of the items in a test were in the first reporting cluster, then 25 percent of the linking items should come from that cluster. The selected linking items were also reviewed by psychometricians to ensure that specific psychometric criteria were met.

After the linking items were approved, assessment specialists populated the rest of the test form. Their first consideration was the strength of the content and the match of each item to a specified content standard. In selecting items, team members also tried to ensure that they included a variety of formats and content and that at least some of the items included graphics for visual interest.

Another consideration was the difficulty of each item. Test developers strived to ensure that there were some easy and some hard items and that there were a number of items in the middle range of difficulty. If items did not meet all content and psychometric criteria, staff reviewed the other available items to determine if there were other selections that could improve the match of the test to all of the requirements. If such a match was not attainable, the content team worked in conjunction with psychometricians and the CDE to determine which combination of items would best serve the needs of the students taking the test. Chapter 3, starting on page 37, contains further information about this process.

Psychometric Criteria

The three goals of STS for RLA test development were as follows:

1. The test must have desired precision of measurement at all ability levels.
2. The test score must be valid and reliable for the intended population and for the various subgroups of test-takers.
3. The test forms must be comparable across years of administration to ensure the generalizability of scores over time.

In order to achieve these goals, a set of rules was developed that outlines the desired psychometric properties of each STS for RLA. These rules are referred to as statistical targets.

Two types of assembly targets were developed for each STS for RLA: the total test target and (reporting) cluster targets. These targets were provided to test developers before a test construction cycle began. The test developers and psychometricians worked together to design the tests to meet these targets.

Primary Statistical Targets

The total test target, or primary statistical targets, used for assembling the STS for RLA forms for the intact forms used in the 2014–15 administration were the test information function (TIF) and an average point-biserial correlation.

The TIF is the sum of the item information function based on the item response theory (IRT) item parameters. When using an IRT model, the target TIF makes it possible to choose items to produce a test that has the desired precision of measurement at all ability levels.

The graphs for each total test are presented in Figure 4.A.1 on page 53. These curves present the target TIF and the projected TIF for the total test at each grade level.

Due to the unique characteristics of the Rasch IRT model, the information curve conditional on each ability level is determined by item difficulty (b -values) alone. In this case, the TIF would, therefore, suffice as the target for conditional test difficulty. Although additional item

difficulty targets are not imperative when the target TIF is used for form construction, the target mean and standard deviation (SD) of item difficulty (b -values) consistent with the TIF were still provided to test development staff to help with the test construction process. The target b -value range approximates a minimum proportion-correct value (p -value) of 0.20 and a maximum p -value of 0.95 for each test.

The point-biserial correlation describes the relationship between student performance on a dichotomously scored item and student performance on the test as a whole. It is used as a measure of how well an item discriminates among test-takers who differ in their ability, and it is related to the overall reliability of the test.

The minimum target value for an item point biserial was set at 0.14 for each test. This value approximates a biserial correlation of 0.20.

Assembly Targets

The target values for the STS for RLA are presented in Table 4.1. These specifications were developed from the analyses of test forms in their original year of administration.

Table 4.1 Statistical Targets for STS for RLA Test Assembly

STS	Target Mean b	Target SD b	P -value Minimum	P -value Maximum	Pt-bis Mean	Pt-Bis Minimum
2 RLA	-0.44	0.91	0.20	0.95	> 0.37	0.14
3 RLA	-0.46	0.88	0.20	0.95	> 0.37	0.14
4 RLA	-0.45	0.71	0.20	0.95	> 0.37	0.14
5 RLA	-0.24	0.65	0.20	0.95	> 0.37	0.14
6 RLA	-0.23	0.72	0.20	0.95	> 0.37	0.14
7 RLA	-0.21	0.80	0.20	0.95	> 0.37	0.14
8 RLA	-0.20	0.74	0.20	0.95	> 0.37	0.14
9 RLA	-0.21	0.77	0.20	0.95	> 0.37	0.14
10 RLA	-0.21	0.79	0.20	0.95	> 0.37	0.14
11 RLA	-0.21	0.70	0.20	0.95	> 0.37	0.14

Target information functions are also used to evaluate the items selected to measure each subscore in the interest of maintaining some consistency in the accuracy of cluster scores across years. Because the clusters include fewer items than the total test, there is always more variability between the target and the information curves constructed for the new form clusters than there is for the total test.

Figure 4.B.1 through Figure 4.B.10, starting on page 55, present the target and projected information curves for the reporting clusters in the administered tests.

Projected Psychometric Properties of the Assembled Tests

In the years when the new forms were developed prior to the 2014–15 administration, Educational Testing Service psychometricians performed a preliminary review of the technical characteristics of the assembled tests. The expected or projected performance of examinees and the overall score reliability were estimated using the item-level statistics available in the California item bank for the selected items. The test reliability was based on Gulliksen's formula (Gulliksen, 1987) for estimating test reliability (r_{xx}) from item p -values and item point-biserial correlations:

$$r_{xx} = \left(\frac{K}{K-1} \right) \left(1 - \frac{\sum_{g=1}^K s_g^2}{\left(\sum_{g=1}^K r_{xg} s_g \right)^2} \right) \quad (4.1)$$

where,

K is the number of items in the test,

s_g^2 is the estimated item variances, i.e., $p_g(1-p_g)$, where p_g is the item p -value for item g ,

r_{xg} is the item point-biserial correlation for item g , and

$r_{xg} s_g$ is the item reliability index.

In addition, estimated test raw score means were calculated by summing the item p -values, and estimated test raw score standard deviations were calculated by summing the item reliability indices. Table 4.A.1 on page 52 presents these summary values by content area and grade.

It should be noted that the projected reliabilities in Table 4.A.1 were based on item p -values and point-biserial correlations. Chapter 8 presents item p -values, point-biserial correlations, and test reliability estimates based on the data from the 2014–15 STS for RLA administration.

Table 4.A.2 on page 52 shows the mean observed statistics of the items on each STS for RLA based on the item-level statistics from the year the form was previously administered. See Table 8.4 on page 111 for the dates of the original administrations. These values can be compared to the target values in Table 4.1.

Rules for Item Sequence and Layout

Because the STS for RLA is primarily passage-dependent, items were sequenced with their associated reading passages. All passages are high interest and were selected and sequenced based on standards and blueprint requirements. Stand-alone items were placed throughout the form, where appropriate.

Reference

Gulliksen, H. (1987). *Theory of mental tests*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Appendix 4.A—Technical Characteristics

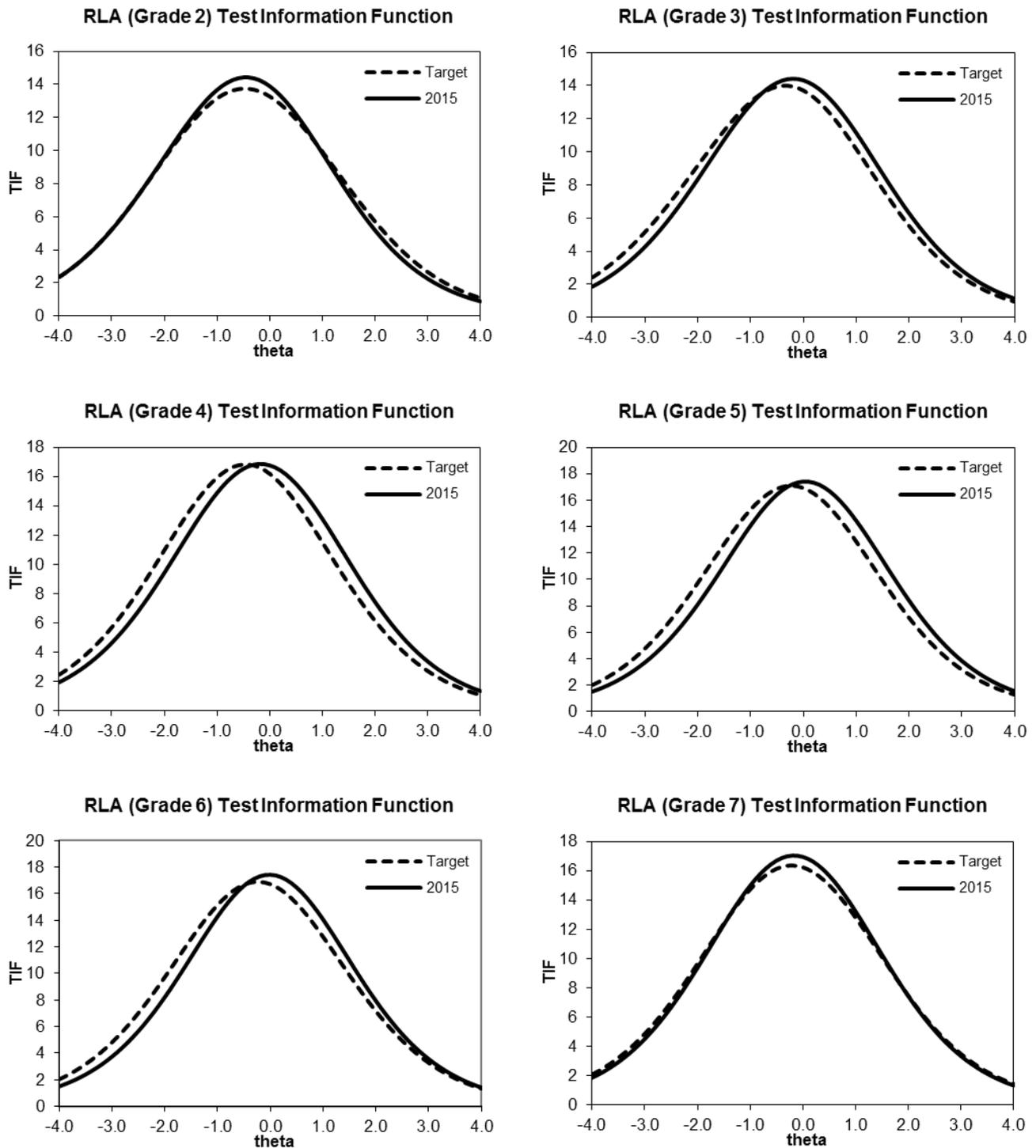
Table 4.A.1 Summary of 2015 STS for RLA Projected Technical Characteristics

STS	Number of Op. Items	Mean Raw Score	Std. Dev. of Raw Scores	Reliability
2 RLA	65	39.46	12.63	0.93
3 RLA	65	36.99	11.63	0.91
4 RLA	75	42.64	14.65	0.93
5 RLA	75	37.69	13.39	0.91
6 RLA	75	38.94	13.39	0.91
7 RLA	75	41.64	13.41	0.92
8 RLA	75	39.75	12.98	0.91
9 RLA	75	40.76	11.25	0.88
10 RLA	75	42.11	11.96	0.89
11 RLA	75	42.30	11.96	0.89

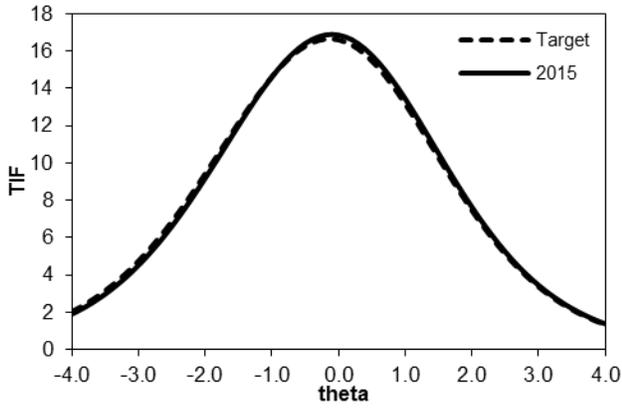
Table 4.A.2 Summary of 2015 STS for RLA Projected Statistical Attributes

STS	Mean b	SD b	Mean <i>p</i> -value	Min <i>p</i> -value	Max <i>p</i> -value	Mean Point Biserial	Min Point Biserial
2 RLA	-0.51	0.76	0.61	0.28	0.92	0.42	0.20
3 RLA	-0.24	0.76	0.57	0.30	0.92	0.38	0.20
4 RLA	-0.19	0.70	0.57	0.28	0.89	0.41	0.19
5 RLA	0.02	0.58	0.50	0.25	0.83	0.37	0.20
6 RLA	-0.02	0.51	0.50	0.24	0.82	0.35	0.16
7 RLA	-0.17	0.65	0.56	0.30	0.79	0.37	0.17
8 RLA	-0.15	0.69	0.53	0.26	0.83	0.35	0.17
9 RLA	-0.20	0.67	0.54	0.27	0.83	0.32	0.14
10 RLA	-0.24	0.74	0.56	0.28	0.88	0.34	0.20
11 RLA	-0.19	0.69	0.56	0.22	0.90	0.34	0.08

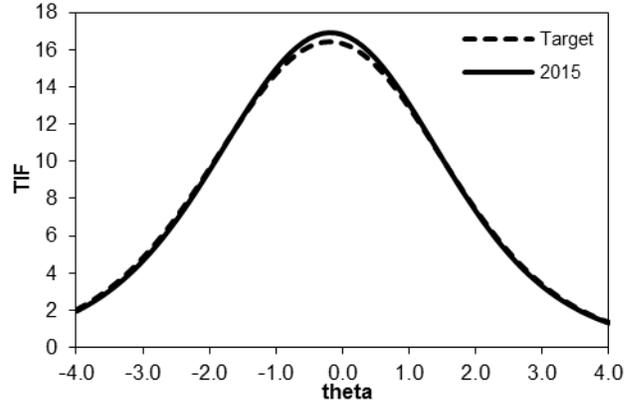
Figure 4.A.1 Plots of Target Information Function and Projected Information for Total Test and Linking Set for RLA



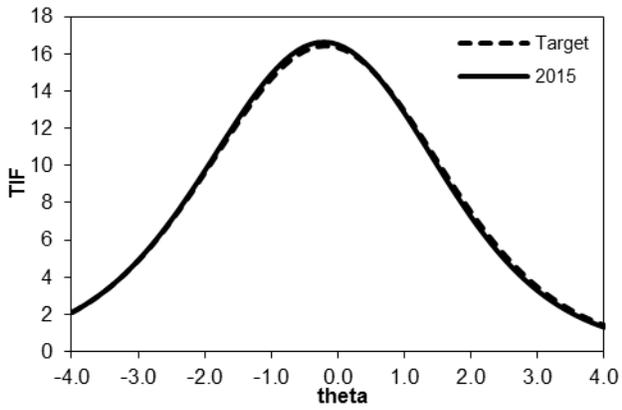
RLA (Grade 8) Test Information Function



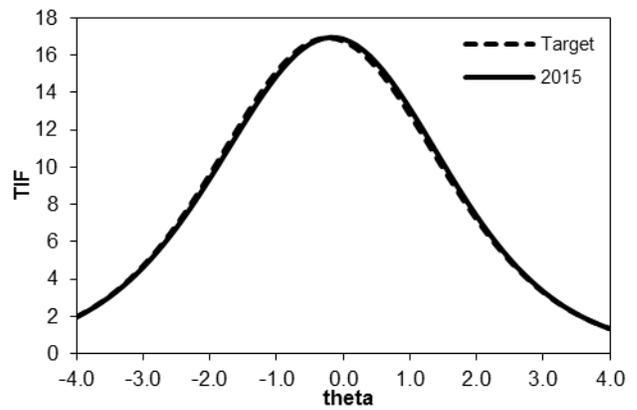
RLA (Grade 9) Test Information Function



RLA (Grade 10) Test Information Function



RLA (Grade 11) Test Information Function



Appendix 4.B—Cluster Targets

Figure 4.B.1 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Two

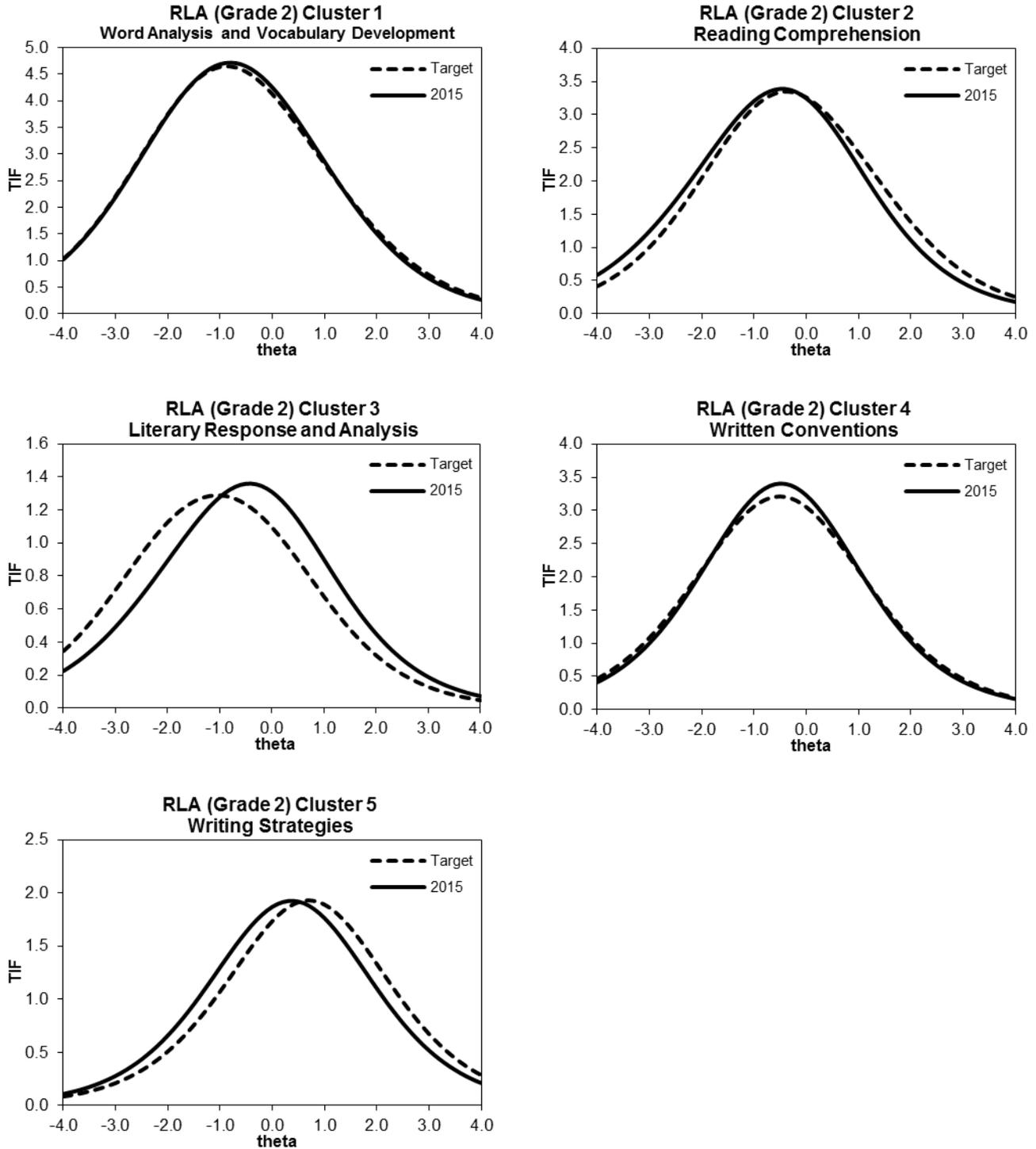


Figure 4.B.2 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Three

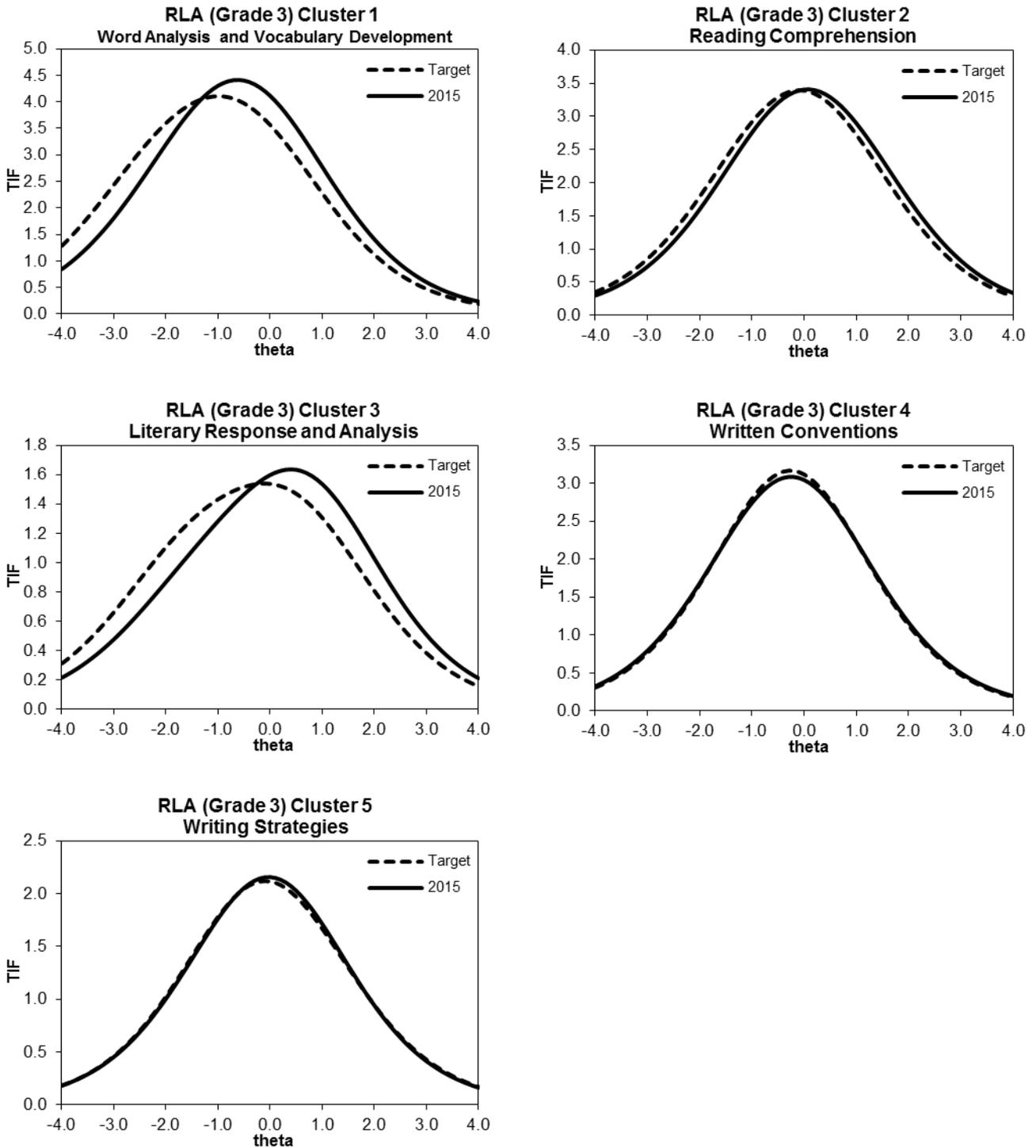


Figure 4.B.3 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Four

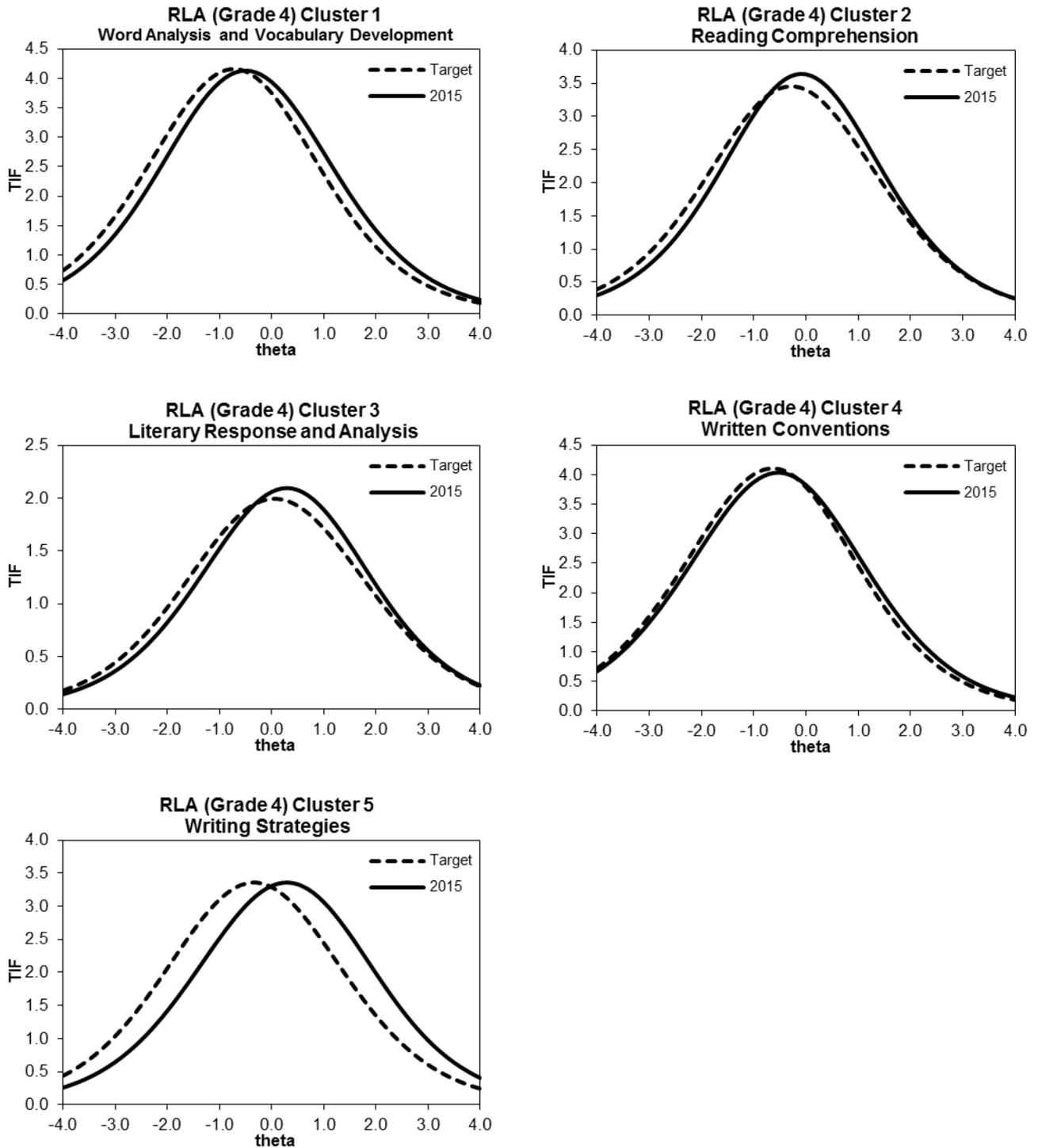


Figure 4.B.4 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Five

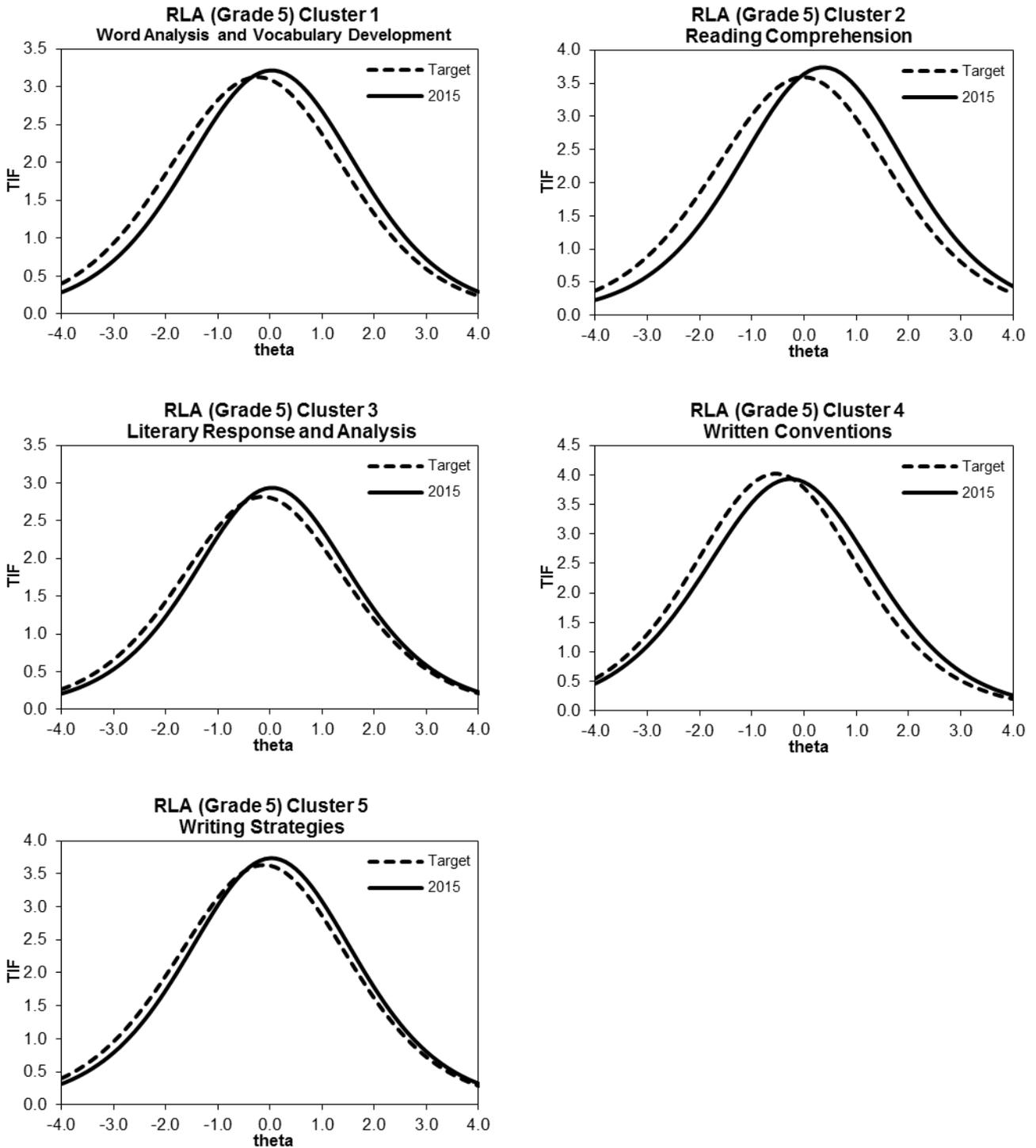


Figure 4.B.5 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Six

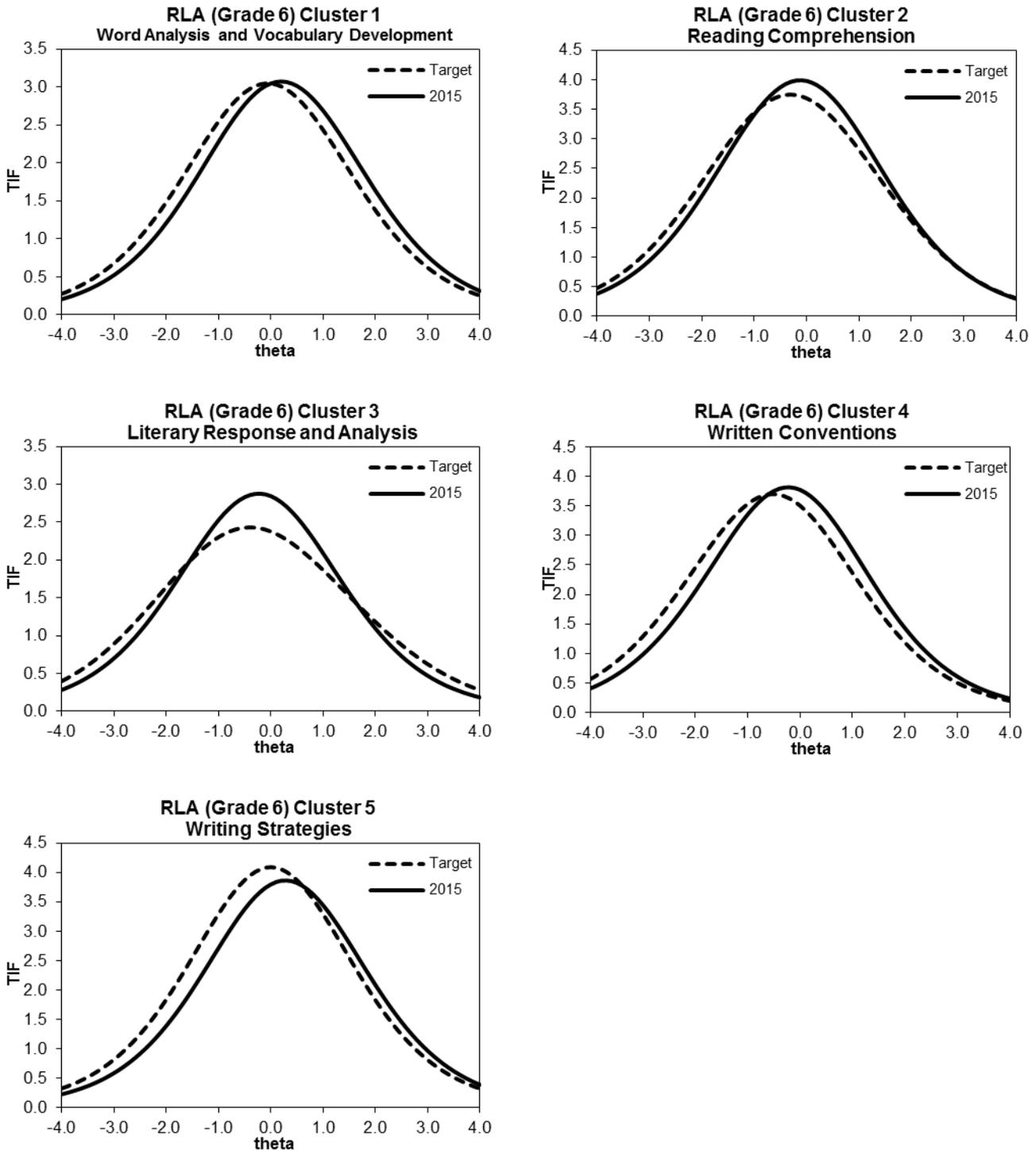


Figure 4.B.6 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Seven

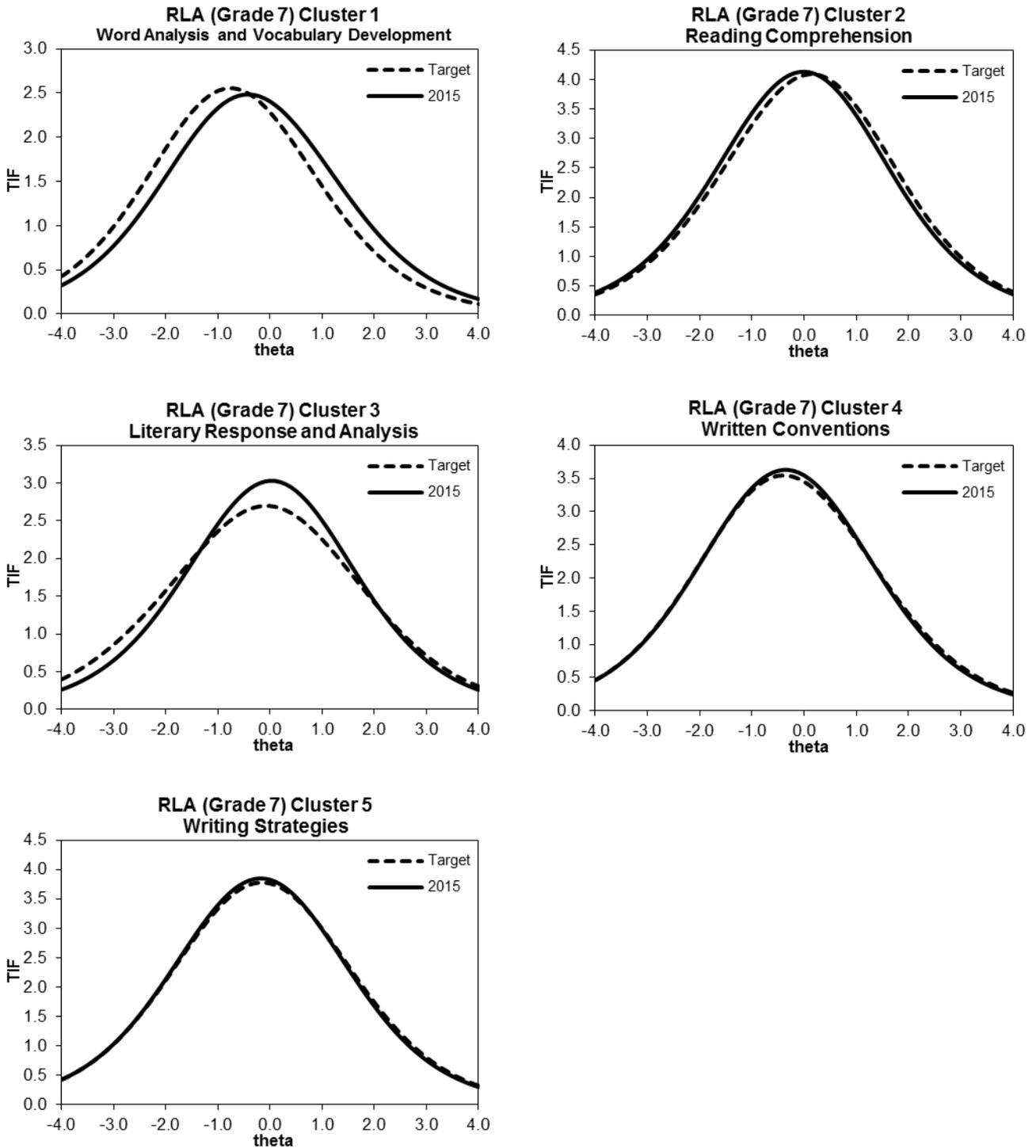


Figure 4.B.7 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Eight

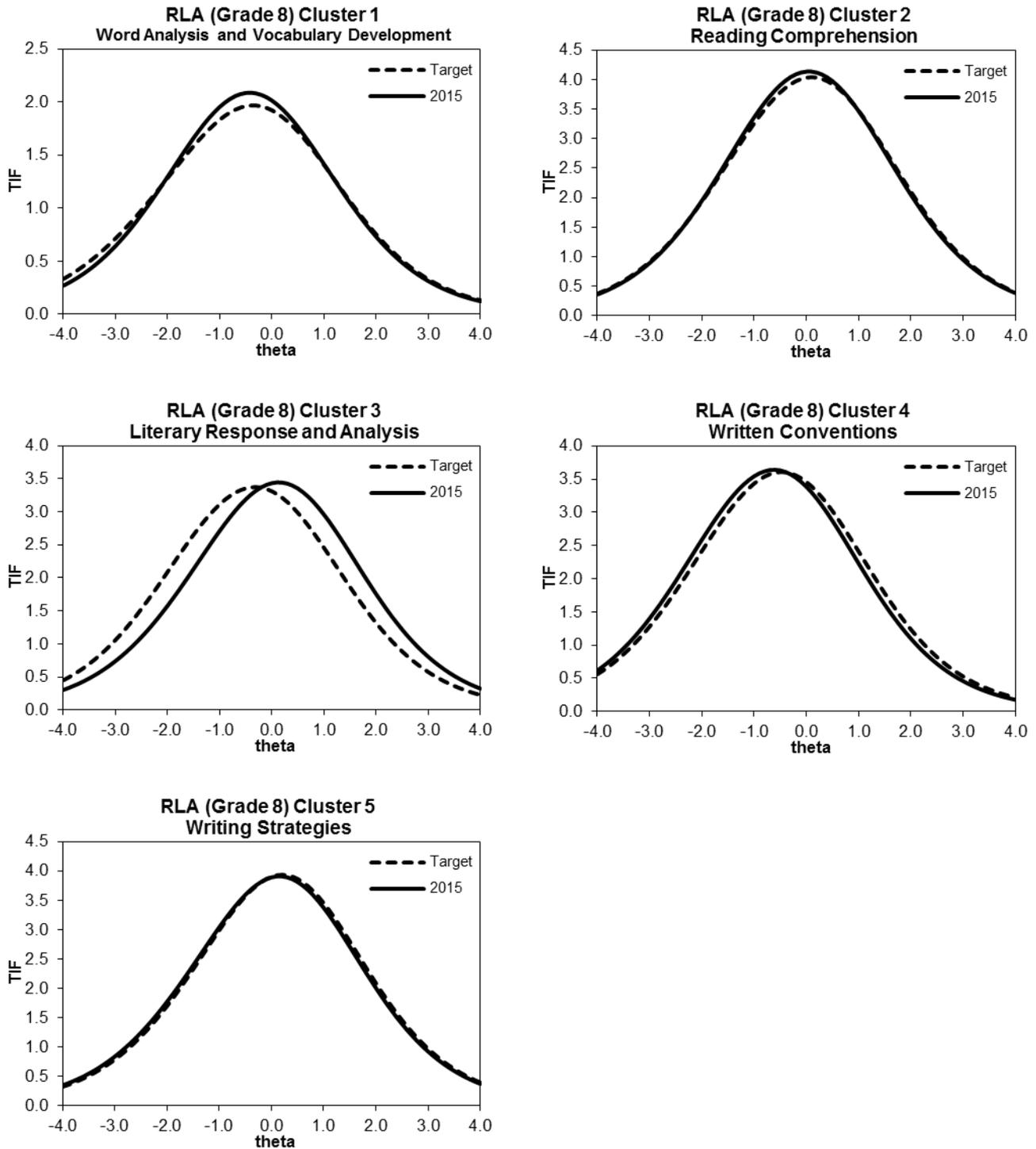


Figure 4.B.8 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Nine

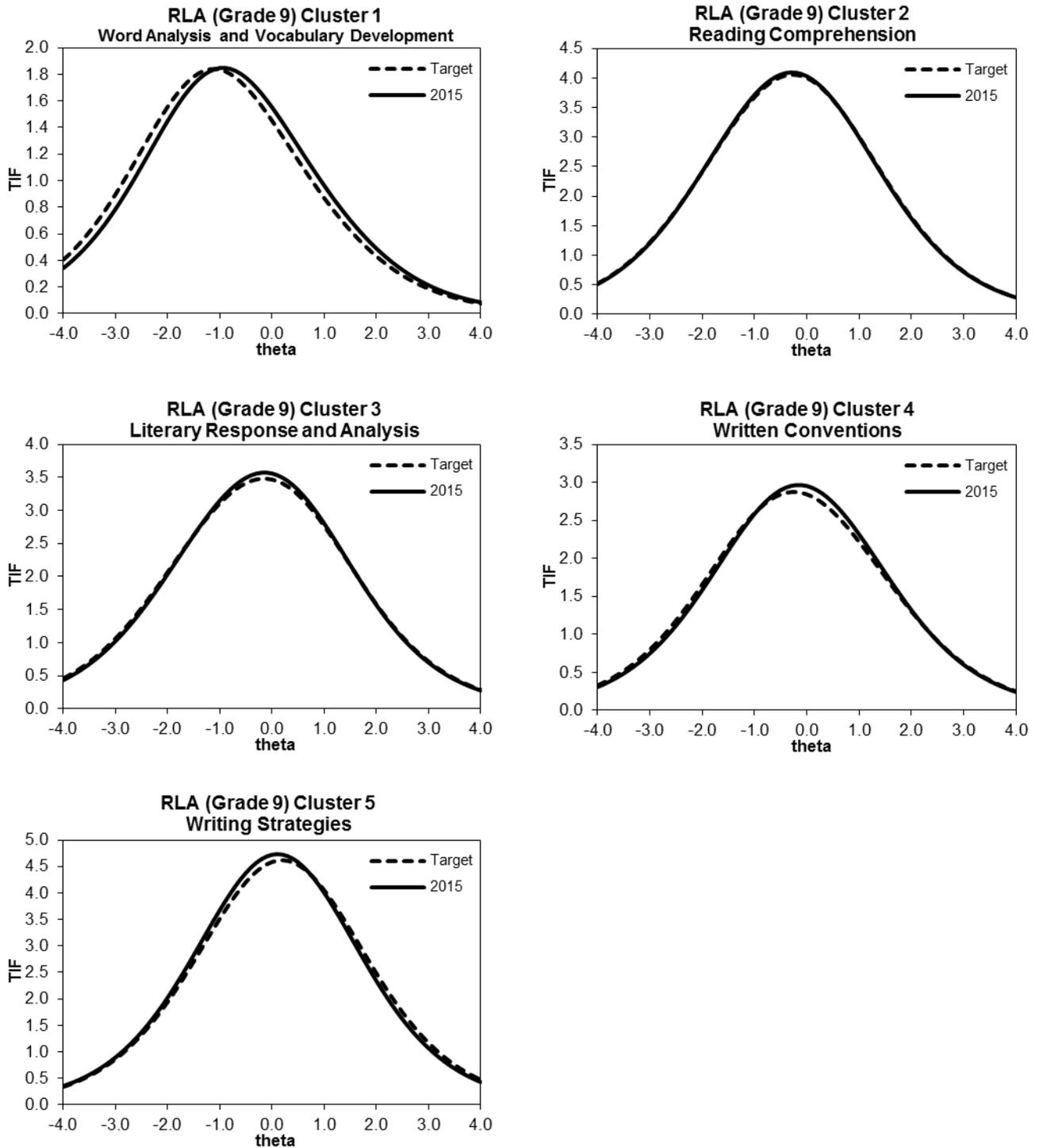


Figure 4.B.9 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Ten

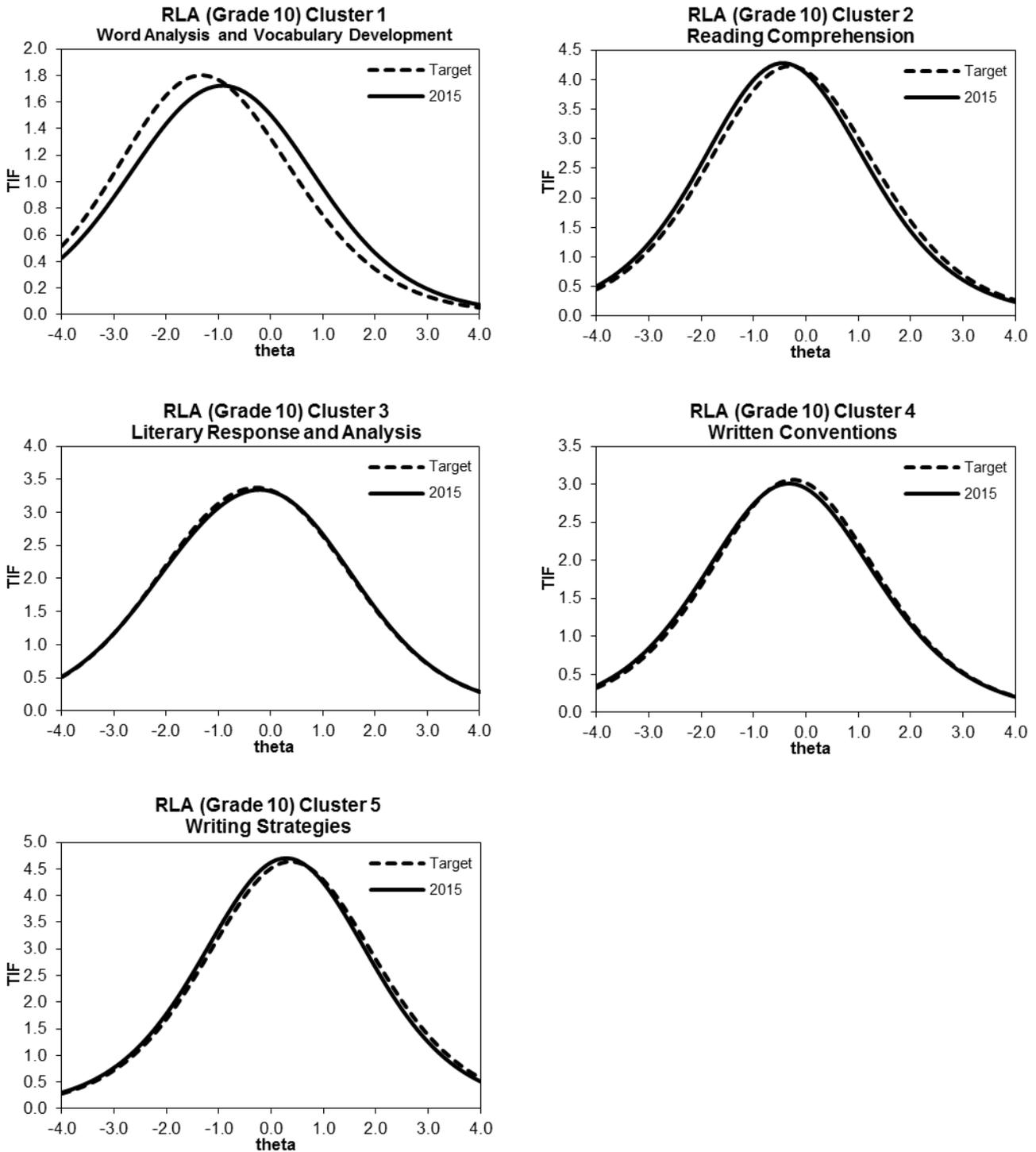
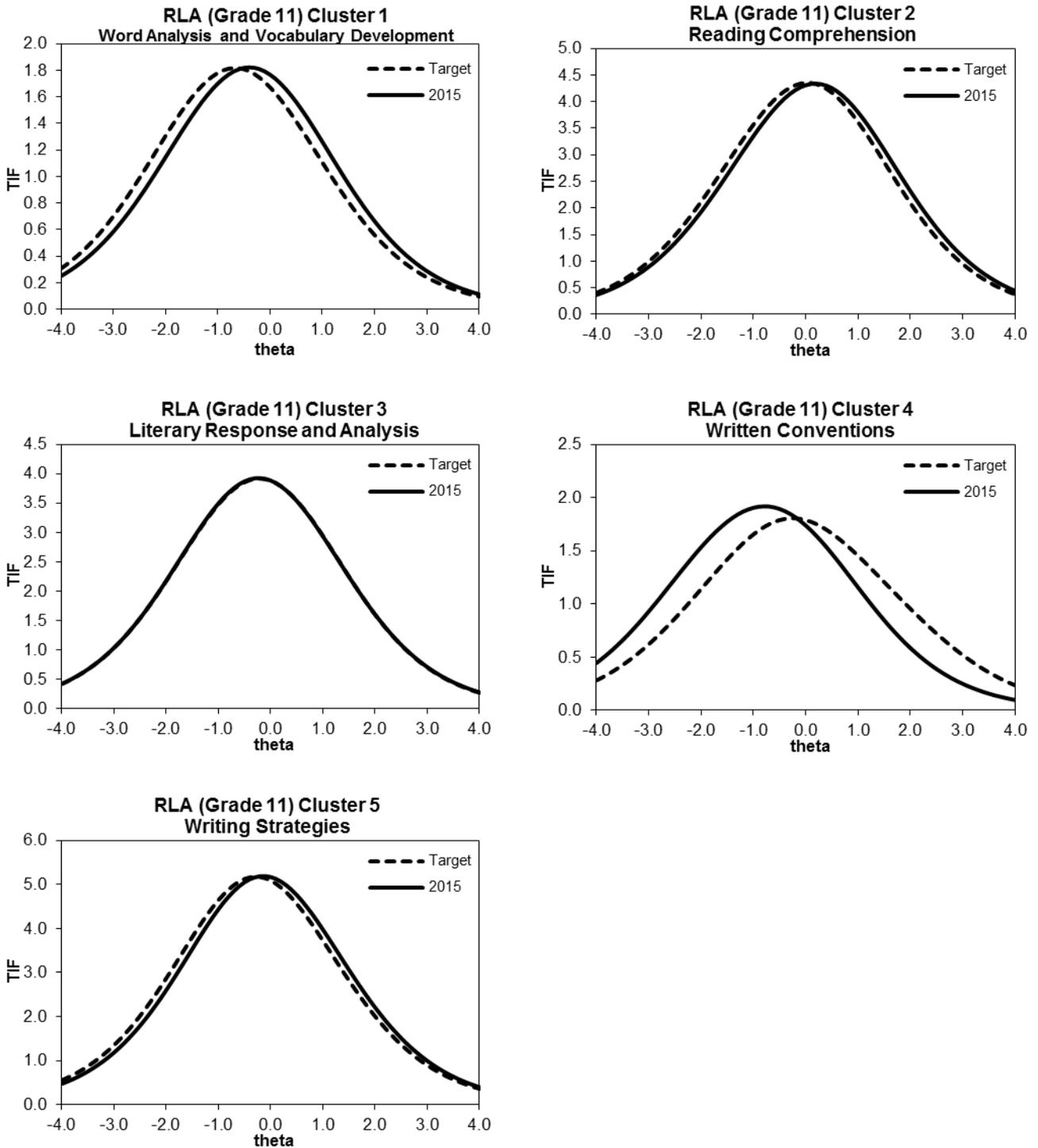


Figure 4.B.10 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Eleven



Chapter 5: Test Administration

Test Security and Confidentiality

All tests within the California Assessment of Student Performance and Progress (CAASPP) System are secure documents. For the Standards-based Tests in Spanish (STS) for Reading/Language Arts (RLA) administration, every person having access to testing materials maintains the security and confidentiality of the tests. Educational Testing Service's (ETS's) Code of Ethics requires that all test information, including tangible materials (such as test booklets), confidential files, processes, and activities are kept secure. ETS has systems in place that maintain tight security for test questions and test results, as well as for student data. To ensure security for all the tests that ETS develops or handles, ETS maintains an Office of Testing Integrity (OTI), which is described in the next section.

ETS's Office of Testing Integrity

The OTI is a division of ETS that provides quality assurance services for all testing programs administered by ETS and resides in the ETS legal department. The Office of Professional Standards Compliance of ETS publishes and maintains *ETS Standards for Quality and Fairness*, which supports the OTI's goals and activities. The purposes of the *ETS Standards for Quality and Fairness* are to help ETS design, develop, and deliver technically sound, fair, and useful products and services, and to help the public and auditors evaluate those products and services.

The OTI's mission is to

- Minimize any testing security violations that can impact the fairness of testing
- Minimize and investigate any security breach
- Report on security activities

The OTI helps prevent misconduct on the part of test-takers and administrators, detects potential misconduct through empirically established indicators, and resolves situations in a fair and balanced way that reflects the laws and professional standards governing the integrity of testing. In its pursuit of enforcing secure practices, ETS, through the OTI, strives to safeguard the various processes involved in a test development and administration cycle. These practices are discussed in detail in the next sections.

Test Development

There was no new item development for the 2014–15 forms. Prior to the 2013–14 administration, during the test development process, ETS staff members consistently adhere to the following established security procedures:

- Only authorized individuals have access to test content at any step during the test development, item review, and data analysis processes.
- Test developers keep all hard-copy test content, computer disk copies, art, film, proofs, and plates in locked storage when not in use.
- ETS shreds working copies of secure content as soon as they are no longer needed during the test development process.
- Test developers take further security measures when testing materials are to be shared outside of ETS; this is achieved by using registered and/or secure mail, using express delivery methods, and actively tracking records of dispatch and receipt of the materials.

Item and Data Review

As mentioned in Chapter 3, Assessment Review Panel (ARP) meetings were not held in 2014–15 because there was no new item development for the 2014–15 STS for RLA forms. However, before the 2014–15 administration, ETS facilitated ARP meetings every year to review all newly developed STS for RLA items and associated statistics. ETS enforced security measures at ARP meetings to protect the integrity of meeting materials using the following guidelines:

- Individuals who participated in the ARPs signed a confidentiality agreement.
- Meeting materials were strictly managed before, during, and after the review meetings.
- Meeting participants were supervised at all times during the meetings.
- Use of electronic devices was prohibited in the meeting rooms.

Item Banking

Once the ARP review was complete, the items were placed in the item bank. ETS then delivered the items to the California Department of Education (CDE) through the California electronic item bank. Subsequent updates to content and statistics associated with items were based on data collected from field testing and the operational use of the items. The latest version of the item is retained in the bank along with the data from every administration that had included the item.

Security of the electronic item banking system is of critical importance. The measures that ETS takes for assuring the security of electronic files include the following:

- Electronic forms of test content, documentation, and item banks are backed up electronically, with the backups kept offsite, to prevent loss from a system breakdown or a natural disaster.
- The offsite backup files are kept in secure storage with access limited to authorized personnel only.
- To prevent unauthorized electronic access to the item bank, state-of-the-art network security measures are used.

ETS routinely maintains many secure electronic systems for both internal and external access. The current electronic item banking application includes a login/password system to provide authorized access to the database or designated portions of the database. In addition, only users authorized to access the specific system query language database are able to use the electronic item banking system. Designated administrators at the CDE and at ETS authorize users to access these electronic systems.

Transfer of Forms and Items to the CDE

ETS shares a secure file transfer protocol (SFTP) site with the CDE. SFTP is a method for reliable and exclusive routing of files. Files reside on a password-protected server that only authorized users may access. On that site, ETS posts Microsoft Word and Excel, Adobe Acrobat PDF, or other document files for the CDE to review. ETS sends a notification e-mail to the CDE to announce that files are posted. Item data are always transmitted in an encrypted format to the SFTP site; test data are never sent via e-mail. The SFTP server is used as a conduit for the transfer of files; secure test data are not stored permanently on the shared SFTP server.

Security of Electronic Files Using a Firewall

A firewall is software that prevents unauthorized entry to files, e-mail, and other organization-specific programs. ETS data exchange and internal e-mail remain within the ETS firewall at all ETS locations, ranging from Princeton, New Jersey, to San Antonio, Texas, to Concord and Sacramento, California.

All electronic applications included in the Test Operations Management System (TOMS) (CDE, 2015a) remain protected by the ETS firewall software at all times. Due to the sensitive nature of the student information processed by TOMS, the firewall plays a significant role in maintaining an assurance of confidentiality in the users of this information.

Printing and Publishing

After items and test forms are approved, the files are sent for printing on a CD using a secure courier system. According to the established procedures, the OTI preapproves all printing vendors before they can work on secured confidential and proprietary testing materials. The printing vendor must submit a completed ETS Printing Plan and a Typesetting Facility Security Plan; both plans document security procedures, access to testing materials, a log of work in progress, personnel procedures, and access to the facilities by the employees and visitors. After reviewing the completed plans, representatives of the OTI visit the printing vendor to conduct an onsite inspection. The printing vendor ships printed test booklets to ETS, which distributes the booklets to local educational agencies (LEAs) in securely packaged boxes.

Test Administration

ETS receives testing materials from printers, packages them, and sends them to LEAs. After testing, the LEAs return materials to ETS for scoring. During these events, ETS takes extraordinary measures to protect the testing materials. ETS uses customized Oracle business applications to verify that inventory controls are in place, from materials receipt to packaging. The reputable carriers used by ETS provide a specialized handling and delivery service that maintains test security and meets the CAASPP System schedule. The carriers provide inside delivery directly to the LEA CAASPP coordinators or authorized recipients of the assessment materials.

Test Delivery

Test security requires accounting for all secure materials before, during, and after each test administration. The LEA CAASPP coordinators are, therefore, required to keep all testing materials in central locked storage except during actual test administration times. CAASPP test site coordinators are responsible for accounting for and returning all secure materials to the LEA CAASPP coordinator, who is responsible for returning them to the Scoring and Processing Center. The following measures are in place to ensure security of CAASPP testing materials:

- LEA CAASPP coordinators are required to sign and submit a “CAASPP Test Security Agreement for LEA CAASPP Coordinators and CAASPP Test Site Coordinators (For all CAASPP assessments, including field tests)” form to the California Technical Assistance Center before ETS can ship any testing materials to the LEA.
- CAASPP test site coordinators have to sign and submit a “CAASPP Test Security Agreement for LEA CAASPP Coordinators and CAASPP Test Site Coordinators (For all CAASPP assessments, including field tests)” form to the LEA CAASPP coordinator before any testing materials can be delivered to the school/test site.

- Anyone having access to the testing materials must sign and submit a “CAASPP Test Security Affidavit for Test Examiners, Proctors, Scribes, and Any Other Persons Having Access to CAASPP Tests (For all CAASPP assessments, including field tests)” form to the CAASPP test site coordinator before receiving access to any testing materials.
- It is the responsibility of each person participating in the CAASPP System to report immediately any violation or suspected violation of test security or confidentiality. The CAASPP test site coordinator is responsible for immediately reporting any security violation to the LEA CAASPP coordinator. The LEA CAASPP coordinator must contact the CDE immediately; the coordinator will be asked to follow up with a written explanation of the violation or suspected violation.

Processing and Scoring

An environment that promotes the security of the test prompts, student responses, data, and employees throughout a project is of utmost concern to ETS. ETS requires the following standard safeguards for security at its sites:

- There is controlled access to the facility.
- No testing materials may leave the facility during the project without the permission of a person or persons designated by the CDE.
- All scoring personnel must sign a nondisclosure and confidentiality form in which they agree not to use or divulge any information concerning tests, scoring guides, or individual student responses.
- All staff must wear ETS identification badges at all times in ETS facilities.

No recording or photographic equipment is allowed in the scoring area without the consent of the CDE.

The completed and scored test booklets and answer documents are stored in secure warehouses. After they are stored, they will not be handled. School and LEA personnel are not allowed to look at a completed test booklet or answer document unless required for transcription or to investigate irregular cases.

All answer documents, test booklets, and other secure testing materials are destroyed after October 31 each year.

Data Management

ETS provides overall security for assessment materials through its limited-access facilities and through its secure data processing capabilities. ETS enforces stringent procedures to prevent unauthorized attempts to access its facilities. Entrances are monitored by security personnel and a computerized badge-reading system is utilized. Upon entering a facility, all ETS employees are required to display identification badges that must be worn at all times while in the facility. Visitors must sign in and out. While they are at the facility, they are assigned a visitor badge and escorted by ETS personnel. Access to the Data Center is further controlled by the computerized badge-reading system that allows entrance only to those employees who possess the proper authorization.

Data, electronic files, test files, programs (source and object), and all associated tables and parameters are maintained in secure network libraries for all systems developed and maintained in a client-server environment. Only authorized software development employees are given access as needed for development, testing, and implementation in a strictly controlled Configuration Management environment.

For mainframe processes, ETS limits and controls access to all data files (test and production), source code, object code, databases, and tables, regulating who is authorized to alter, update, or even read the files. All attempts to access files on the mainframe by unauthorized users are logged and monitored. In addition, ETS controls versions of the software and data files. ChangeMan provides another level of security to place the correct tested version of code into production. Unapproved changes are not implemented without prior review and approval.

Statistical Analysis

The Information Technology (IT) department at ETS loads data files into a database. The Data Quality Services group at ETS extracts the data from the database and performs quality control procedures before passing files to the ETS Statistical Analysis group. The Statistical Analysis group keeps the files on secure servers and adheres to the ETS Code of Ethics and the ETS Information Protection Policies to prevent any unauthorized access.

Reporting and Posting Results

After statistical analysis has been completed on student data, the following deliverables are produced:

- Printed Student Score Reports are produced and shipped to the designated LEA for distribution
- PDFs of Student Score Reports available through TOMS
- A file of individual student results—available for download from TOMS—that shows students' scale scores and performance levels
- A file of aggregated student results available for download through TOMS
- Encrypted files of summary results (sent to the CDE by means of SFTP) (Any summary results that have fewer than 11 students are not reported.)
- Item-level statistics based on the results, which are entered into the item bank

Student Confidentiality

To meet Elementary and Secondary Education Act and state requirements, LEAs must collect demographic data about students. This includes information about students' ethnicity, parent education, disabilities, whether the student qualifies for the National School Lunch Program, and so forth (CDE, 2015b). ETS takes precautions to prevent any of this information from becoming public or being used for anything other than testing purposes. These procedures are applied to all documents in which these student demographic data may appear, including in Pre-ID files and reports.

Student Test Results

ETS also has security measures to protect files and reports that show students' scores and performance levels. ETS is committed to safeguarding the information in its possession from unauthorized access, disclosure, modification, or destruction. ETS has strict information security policies in place to protect the confidentiality of ETS and client data. ETS staff access to production databases is limited to personnel with a business need to access the data. User IDs for production systems must be person-specific or for systems use only.

ETS has implemented network controls for routers, gateways, switches, firewalls, network tier management, and network connectivity. Routers, gateways, and switches represent points of access between networks. However, these do not contain mass storage or

represent points of vulnerability, particularly to unauthorized access or denial of service. Routers, switches, firewalls, and gateways may possess little in the way of logical access.

ETS has many facilities and procedures that protect computer files. Facilities, policies, software, and procedures such as firewalls, intrusion detection, and virus control are in place to provide for physical security, data security, and disaster recovery. ETS is certified in the BS 25999-2 standard for business continuity and conducts disaster recovery exercises annually. ETS routinely backs up its data to either disk through deduplication or to tape, both of which are stored off site.

Access to the ETS Computer Processing Center is controlled by employee and visitor identification badges. The Center is secured by doors that can only be unlocked by the badges of personnel who have functional responsibilities within its secure perimeter. Authorized personnel accompany visitors to the Processing Center at all times. Extensive smoke detection and alarm systems, as well as a pre-action fire-control system, are installed in the Center.

ETS protects individual students' results on both electronic files and paper reports during the following events:

- Scoring
- Transfer of scores by means of secure data exchange
- Reporting
- Analysis and reporting of erasure marks
- Posting of aggregate data
- Storage

In addition to protecting the confidentiality of testing materials, ETS's Code of Ethics further prohibits ETS employees from financial misuse, conflicts of interest, and unauthorized appropriation of ETS's property and resources. Specific rules are also given to ETS employees and their immediate families who may take a test developed by ETS, such as a CAASPP examination. The ETS Office of Testing Integrity verifies that these standards are followed throughout ETS. It does this, in part, by conducting periodic onsite security audits of departments, with follow-up reports containing recommendations for improvement.

Procedures to Maintain Standardization

The STS for RLA processes are designed so that the tests are administered and scored in a standardized manner.

ETS employs personnel who facilitate various processes involved in the standardization of an administration cycle and takes all necessary measures to ensure the standardization of the STS for RLA, as described in this section.

Test Administrators

The STS for RLA are administered in conjunction with the other tests that comprise the CAASPP System. The responsibilities for LEA and test site staff members are included in the *CAASPP Paper-Pencil Testing Test Administration Manual* (CDE, 2015b). This manual is described in the next section.

The staff members centrally involved in the test administration are as follows:

LEA CAASPP Coordinator

Each LEA designates an LEA CAASPP coordinator who is responsible for ensuring the proper and consistent administration of the CAASPP tests. LEAs include public school districts, statewide benefit charter schools, state board-authorized charter schools, county office of education programs, and charter schools testing independently from their home district.

LEA CAASPP coordinators are also responsible for securing testing materials upon receipt, distributing testing materials to schools, tracking the materials, training and answering questions from LEA staff and CAASPP test site coordinators, reporting any testing irregularities or security breaches to the CDE, receiving scorable and nonscorable materials from schools after an administration, and returning the materials to the CAASPP contractor for processing.

CAASPP Test Site Coordinator

The superintendent of the school district or the LEA CAASPP coordinator designates a CAASPP test site coordinator at each test site from among the employees of the LEA. (*California Code of Regulations, Title 5 [5 CCR], Section 858[a]*)

CAASPP test site coordinators are responsible for making sure that the school has the proper testing materials, distributing testing materials within a school, securing materials before, during, and after the administration period, answering questions from test examiners, preparing and packaging materials to be returned to the LEA after testing, and returning the materials to the LEA. (CDE, 2015b)

Test Administrator

The STS for RLA are administered by test administrators who may be assisted by test proctors and scribes. A test administrator is an employee of an LEA or an employee of a nonpublic, nonsectarian school (NPS) who has been trained to administer the tests, has signed a CAASPP Test Security Affidavit, and is bilingual in English and Spanish. Test administrators must follow the directions in the *California Standards-based Tests in Spanish Directions for Administration (DFA)* (CDE, 2015c) exactly.

Test Proctor

A test proctor is an employee of an LEA or a person, assigned by an NPS to implement the individualized education program (IEP) of a student, who has received training designed to prepare the proctor to assist the test examiner in the administration of tests within the CAASPP System (5 CCR Section 850[y]). Test proctors must sign CAASPP Test Security Affidavits (5 CCR Section 859 [c]).

Scribe

A scribe is an employee of an LEA or a person, assigned by an NPS to implement the IEP of a student, who is required to transcribe a student's responses to the format required by the test. A student's parent or guardian is not eligible to serve as the student's scribe (5 CCR Section 850[s]). Scribes must sign CAASPP Test Security Affidavits (5 CCR Section 859[c]).

Directions for Administration (DFAs)

STS for RLA *Directions for Administration (DFAs)* are manuals used by test administrators to administer the STS for RLA to students (CDE, 2015c). Test administrators must follow all directions and guidelines and read, word-for-word, the instructions to students in "SAY" boxes to ensure test standardization.

CAASPP Paper-Pencil Testing Test Administration Manual

Test administration procedures are to be followed exactly so that all students have an equal opportunity to demonstrate their academic achievement. The *CAASPP Paper-Pencil Testing Test Administration Manual* contributes to this goal by providing information about the responsibilities of LEA CAASPP and CAASPP test site coordinators, as well as those of the other staff involved in the administration cycle (CDE, 2015b). However, the manual is not intended as a substitute for *5 CCR*, or to detail all of the coordinator's responsibilities.

Test Operations Management System Manuals

TOMS is a series of secure, Web-based modules that allow LEA CAASPP coordinators to set up test administrations and ensure test sites order materials. Every module has its own user manual with detailed instructions on how to use TOMS. The TOMS modules used to manage paper-pencil test processes are as follows:

- **Test Administration Setup**—This module allows LEAs to determine and calculate dates for scheduling test administrations for LEAs, verify contact information for those LEAs, and request Pre-ID labels. (CDE, 2015d)
- **Student Paper-Pencil Test Registration**—This module allows LEAs to assign paper-pencil science tests to students in grades five, eight, and ten. (CDE, 2015e)
- **Set Condition Codes**—This module allows LEA CAASPP coordinators and CAASPP test site coordinators to apply condition codes (to note that a student was absent during testing, for example) to student records.

Test Booklets

For each grade-level test, multiple versions of test booklets are administered. The versions differ only in terms of the field-test items they contain. In grades three through eleven, these versions are spiraled—comingled—and packaged consecutively and are distributed at the student level; that is, each classroom or group of test-takers receives at least one of each version of the test. The grade two STS for RLA versions are not spiraled; instead, versions are assigned by school.

The test booklets, along with answer documents and other supporting materials, are packaged by school or group. All materials are sent to the LEA CAASPP coordinator for proper distribution within the LEA. Special formats of test booklets are also available for test-takers who require accommodations to participate in testing. These special formats include large-print and braille testing materials.

Universal Tools, Designated Supports, and Accommodations for Students with Disabilities

All public school students participate in the CAASPP System, including students with disabilities and English learners. ETS policy states that reasonable testing accommodations be provided to candidates with documented disabilities that are identified in the Americans with Disabilities Act (ADA). The ADA mandates that test accommodations be individualized, meaning that no single type of test accommodation may be adequate or appropriate for all individuals with any given type of disability. The ADA authorizes that test-takers with disabilities may be tested under standard conditions if ETS determines that only minor adjustments to the testing environment are required (e.g., wheelchair access, large-print test book, a sign language interpreter for spoken directions).

Identification

Most students with disabilities take the STS for RLA under standard conditions. However, some students with disabilities may need assistance when taking the STS for RLA. This assistance takes the form of universal tools, designated supports, and accommodations (see Appendix 2.C on page 21 in Chapter 2 for details). During the test, these students may use the special services specified in their IEP or Section 504 plan. If students use universal tools, designated supports, and/or accommodations for the STS for RLA, test examiners are responsible for marking the universal tools, designated supports, and/or accommodations used on the students' test booklets or answer documents.

Scoring

The purpose of universal tools, designated supports, and accommodations in testing is to allow *all* students the opportunity to demonstrate what they know and what they are able to do, rather than give students using them an advantage over other students or artificially inflate their scores. Universal tools, designated supports, and accommodations minimize or remove the barriers that could otherwise prevent students from generating results that reflect their achievement in the content area.

Testing Incidents

Testing incidents—breaches and irregularities—are circumstances that may compromise the reliability and validity of test results

The LEA CAASPP coordinator is responsible for immediately notifying the CDE of any irregularities or breaches that occur before, during, or after testing. The test examiner is responsible for immediately notifying the LEA CAASPP coordinator of any security breaches or testing irregularities that occur in the administration of the test. Once the LEA CAASPP coordinator and the CDE have determined that an irregularity or breach has occurred, the CDE instructs the LEA CAASPP coordinator on how and where to identify the irregularity or breach on the student answer document. The information and procedures to assist in identifying incidents and notifying the CDE are provided in the *CAASPP Paper-Pencil Testing Test Administration Manual* (CDE, 2015b).

Social Media Security Breaches

Social media security breaches are exposures of test questions and testing materials through social media Web sites. These security breaches raise serious concerns that require comprehensive investigation and additional statistical analyses. In recognizing the importance of and the need to provide valid and reliable results to the state, LEAs, and schools, both the CDE and ETS take every precaution necessary, including extensive statistical analyses, to ensure that all test results maintain the highest levels of psychometric integrity.

There were no social media security breaches associated with the STS for RLA in 2014–15 that required any item to be withheld from scoring.

Testing Improperities

A testing impropriety is any event that occurs before, during, or after test administrations that does not conform to the instructions stated in the *DFAs* (CDE, 2015c) and the *CAASPP Paper-Pencil Testing Test Administration Manual* (CDE, 2015b). These events include test administration errors, disruptions, and student cheating. Testing improprieties generally do not affect test results and are not reported to the CDE or the CAASPP System testing contractor. The CAASPP test site coordinator should immediately notify the LEA CAASPP coordinator of any testing improprieties that occur. It is recommended by the CDE that LEAs and schools maintain records of testing improprieties.

References

- California Department of Education. (2015a). *2015 Test Management System*. Sacramento, CA. <http://www.caaspp.org/administration/toms/>
- California Department of Education. (2015b). *2015 CAASPP LEA and test site coordinator manual*. Sacramento, CA. Retrieved from http://www.caaspp.org/rsc/pdfs/CAASPP.coord_man.2015.pdf
- California Department of Education. (2015c). *2015 California Standards-based Tests in Spanish directions for administration*. Sacramento, CA. Retrieved from http://www.caaspp.org/rsc/pdfs/STS.grade-5_dfa.2015.pdf
- California Department of Education. (2015d). *California Assessment of Student Performance and Progress Test Operations Management System: Test administration setup guide*. Sacramento, CA. Retrieved from http://www.caaspp.org/rsc/pdfs/CAASPP.test_admin_setup.2015.pdf
- California Department of Education. (2015e). *2015 California Assessment of Student Performance and Progress Test Operations Management System: Student paper-pencil test registration user guide*. Sacramento, CA. Retrieved from <http://www.caaspp.org/rsc/pdfs/CAASPP.ppt-registration.2015.pdf>

Chapter 6: Performance Standards

Background

The Standards-based Tests in Spanish (STS) were introduced to California’s standardized testing program in stages, starting with the lower grades in 2006. Performance standards for each new test were developed after the introductory year for operational use in subsequent administrations. The STS for reading/language arts (RLA) and mathematics in grades two through four were established in 2006. For each of these tests, the performance standards were developed in February 2009 and adopted by the State Board of Education (SBE) for their 2009 operational administration.

In spring 2008, the STS for RLA and mathematics were introduced in grades five through seven. The performance standards for those tests were developed in October 2009 and adopted by the SBE for the 2010 operational administration.

The STS for RLA in grades eight through eleven and end-of-course mathematics were introduced in spring 2009. The performance standards for those tests were developed in November 2011 and adopted by the SBE in July 2012 for use starting with the 2013 operational administration.

The performance standards for the STS were defined by the SBE as far below basic, below basic, basic, proficient, and advanced. Performance standards are developed from a general description of the performance level (policy-level descriptors) and competencies lists, which operationally define each level. Cut scores numerically define the performance levels.

In 2014–15, the STS for RLA were administered to eligible students in grades two through eleven. Consequently, the performance standards for the grades and subjects were applied to the scores of students.

California employed carefully designed standard-setting procedures to facilitate the development of performance standards for each STS. The standard-setting method used for the STS is the Bookmark method (Mitzel, Lewis, Patz, & Green, 2001). These processes are described in the sections that follow.

Standard-Setting Procedure

The process of standard setting is designed to identify a “cut score” or minimum test score that is required to qualify a student for each performance level. The process generally requires a panel of subject-matter experts and others with relevant perspectives (for example, teachers or school administrators) be assembled. The panelists for the STS standard settings were selected based on the following characteristics:

- Familiarity with the subject matter assessed
- Familiarity with students in the respective grade levels
- Experience with English learners (ELs)
- Familiarity with the California content standards
- An understanding of the STS
- An appreciation of the consequences of setting these cut scores

All panelists were bilingual and biliterate in Spanish and English, were recruited from diverse geographic regions, and were from different gender and major racial/ethnic subgroups to be representative of the educators of the state’s STS-eligible students (Educational Testing Service [ETS], 2009, 2010, 2011).

For each test, three cut scores were developed in order to differentiate four of the five performance levels: below basic, basic, proficient, and advanced. Far below basic was defined as chance-level performance.

The standard-setting processes implemented for the STS required panelists to follow these steps, which include training and practice prior to making judgments:

1. Prior to attending the workshop, all panelists received a pre-workshop assignment. The task was to review, on their own, the content standards upon which the test items are based and take notes on their own expectations in the content area. This allowed the panelists to understand how their perceptions may relate to the complexity of the content standards.
2. At the start of the workshop, panelists received training, which included the purpose of standard setting and their role in the work, the meaning of a “cut score” and “impact data,” and specific training and practice in the Bookmark method. Impact data included the percentage of examinees assessed in a previous administration of the test that would fall into each level, given the panelists’ judgments of cut scores.
3. Panelists became familiar with the difficulty level of the items by taking the actual test and then assessing and discussing the demands of the test items.
4. Panelists reviewed the draft list of competencies as a group, noting the increasing demands of each subsequent level. In this step, they began to visualize the knowledge and skills of students in each performance level.
5. Panelists identified characteristics of a “borderline” test-taker or “target student.” This student is defined as one who possesses just enough knowledge of the content to move over the border separating a performance level from the performance level below it.
6. After training in the method was complete and confirmed through an evaluation questionnaire, panelists made individual judgments. Working in small groups, they discussed feedback related to other panelists’ judgments and feedback based on student performance data (impact data). Panelists could revise their judgments during the process if they wished.
7. The final recommended cut scores were based on the median of panelists’ judgment scores at the end of three rounds (in the Bookmark method, the panel recommendation is calculated by taking the median of the small group [table] medians). For the STS for RLA, the cut scores recommended by the panelists and the recommendation of the State Superintendent of Public Instruction were presented for public comment at regional public hearings. Comments and recommendations were then presented to the SBE for adoption.

Development of Competencies Lists

Prior to the STS standard-setting workshop, ETS facilitated a meeting in which a subset of the standard-setting panelists was assembled to develop lists of competencies based on the California content standards and policy-level descriptors. For each content area, one panel of educators was assembled for each grade to identify and discuss the competencies

required of students taking the STS for each performance level (below basic, basic, proficient, and advanced). The lists were used to facilitate the discussion and construction of the target student definitions during the standard-setting workshop.

Standard-Setting Methodology

Bookmark Method

The Bookmark method for setting cut scores was introduced in 1999 and has been used widely across the United States (Lewis, et al., 1999; Mitzel, et al., 2001). In California, the Bookmark method was used in standard settings for most of the CAASPP paper-pencil tests.

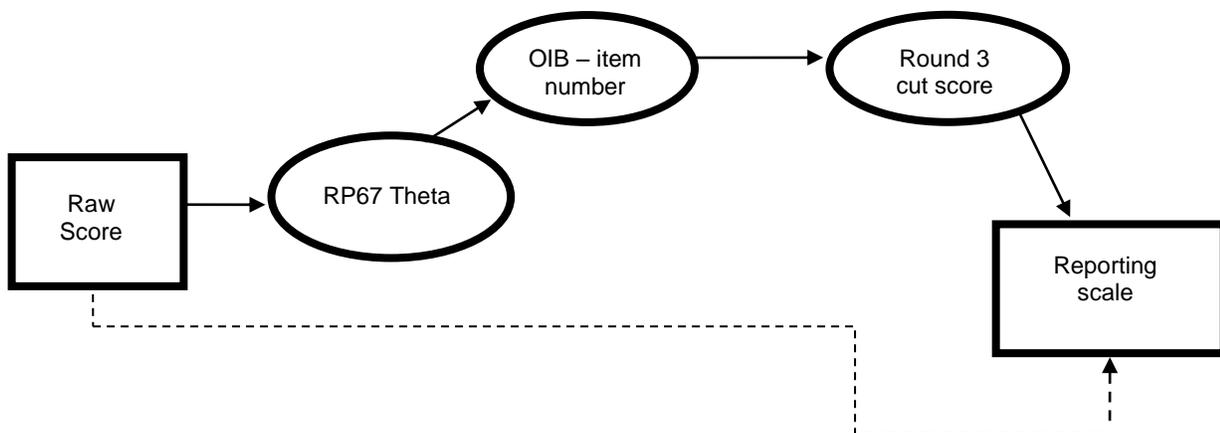
The Bookmark method is an item-mapping procedure in which panelists consider content covered by items in a specially constructed book where items are ordered from easiest to hardest based on operational student performance data from a previous test administration. The “item map,” which accompanies the ordered item booklet (OIB), includes information on the content measured by each operational test question, information about each question’s difficulty, the correct answer for each question, and where each question was located in the test booklet before the questions were reordered by difficulty.

Panelists are asked to place a bookmark in the OIB to demarcate each performance level. The bookmarks are placed with the assumption that the borderline students will perform successfully at a given performance level with a probability of at least 0.67. Conversely, these students are expected to perform successfully on the items after the bookmark with a probability of less than 0.67 (Huynh, 1998).

In this method, the panelists’ cut-score recommendations are presented in the metric of the OIB and are derived by obtaining the median of the corresponding bookmarks placed for each performance level across panelists.

Each item location corresponds to a value of theta, based on a response probability of 0.67 (RP67 Theta), which maps back to a raw score on the test form. Figure 6.1 below may best illustrate the relationship among the various metrics used when the Bookmark method is applied. The solid lines represent steps in the standard-setting process described above; the dotted line represents the scaling described in the next section.

Figure 6.1 Bookmark Standard-setting Process for the STS



Results

The cut scores obtained as a result of the standard-setting process were on the item response theory (IRT) scale; each recommended cut score was associated with a theta value in the OIB. This RP67 Theta has a corresponding number-correct or raw score for the test form upon which standards were set; the scores were then translated to a score scale that ranges between 150 and 600.

The cut score for the basic performance level was set to 300 for every grade and content area; this means that a student must earn a score of 300 or higher to achieve a basic classification. The cut score for the proficient performance level was set to 350 for every grade and content area; this means that a student must earn a score of 350 or higher to achieve a proficient classification.

The cut scores for the other performance levels were derived using procedures based on IRT and usually vary by grade and content area. Each raw cut score for a given test was mapped to an IRT *theta* (θ) using the test characteristic function or curve and then transformed to the scale-score metric using the following equation:

$$\text{Scale Cut Score} = (350 - \theta_{\text{proficient}} \times \left(\frac{350 - 300}{\theta_{\text{proficient}} - \theta_{\text{basic}}} \right)) + \left(\frac{350 - 300}{\theta_{\text{proficient}} - \theta_{\text{basic}}} \right) \times \theta_{\text{cut-score}} \quad (6.1)$$

where,

$\theta_{\text{cut-score}}$ represents the student ability at cut scores for performance levels other than proficient or basic, e.g., below basic or advanced,

$\theta_{\text{proficient}}$ represents the theta corresponding to the cut score for proficient, and

θ_{basic} represents the theta corresponding to the cut score for basic.

Please note that an IRT test characteristic function or curve is the sum of item characteristic curves (ICC), where an ICC represents the probability of correctly responding to an item conditioned on examinee ability.

The scale-score ranges for each performance level are presented in Table 2.1 on page 16. The cut score for each performance level is the lower bound of each scale-score range. The scale-score ranges do not change from year to year. Once established, they remain unchanged from administration to administration until such time that new performance standards are adopted.

Table 7.2 on page 84 in Chapter 7 presents the percentages of examinees meeting each performance level for the 2014–15 administration.

References

- Educational Testing Service. (2009). *Technical report on the standard setting workshop for the California Standards-based Tests in Spanish: RLA grades two through four, mathematics grades two through four, March 20, 2009* (California Department of Education Contract Number 5417). Princeton, NJ: Author.
- Educational Testing Service. (2010). *Technical report on the standard setting workshop for the California Standards-based Tests in Spanish: RLA grades five through seven, mathematics grades five through seven, January 15, 2010* (California Department of Education Contract Number 5417). Princeton, NJ: Author.
- Educational Testing Service. (2011). *Technical report on the standard setting workshop for the California Standards-based Tests in Spanish: RLA grades eight through eleven, Algebra I, and Geometry, December 28, 2011* (California Department of Education Contract Number 5417). Princeton, NJ: Author.
- Huynh, H. (1998). On score locations of binary and partial credit items and their applications to item mapping and criterion-referenced interpretation. *Journal of Educational and Behavioral Statistics*, 23(19), 35–56.
- Lewis, D. M., Green, D. R., Mitzel, H. C., Baum, K., & Patz, R. J. (1999). *The bookmark standard setting procedure: Methodology and recent implications*. Manuscript under review.
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives*, (pp. 249–81). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Chapter 7: Scoring and Reporting

Educational Testing Service (ETS) conforms to high standards of quality and fairness (ETS, 2002) when scoring tests and reporting scores. These standards dictate that ETS provides accurate and understandable assessment results to the intended recipients. It is also ETS's mission to provide appropriate guidelines for score interpretation and cautions about the limitations in the meaning and use of the test scores. Finally, ETS conducts analyses needed to ensure that the assessments are equitable for various groups of test-takers.

Procedures for Maintaining and Retrieving Individual Scores

Items for all the Standards-based Tests in Spanish (STS) for Reading/Language Arts (RLA) are multiple choice. Students are presented with a question and asked to select the correct answer from among four possible choices. In grades two and three, students mark their answer choices in the test booklet. In the other grades, students mark their answer choices in an answer document. All questions are machine scored.

In the 2014–15 administration, because the raw-score-to-scale-score conversion tables were developed before tests were administered using pre-equating, preliminary individual student results were available for download prior to the printing of paper reports. This electronic reporting was made possible through the Online Reporting System.

In order to score and report STS for RLA results, ETS follows an established set of written procedures. The specifications for these procedures are presented in the next sections.

Scoring and Reporting Specifications

ETS develops standardized scoring procedures and specifications so that test materials are processed and scored accurately. These documents include the following:

- **Scoring Rules**—Describes the following:
 - the rules for how and when scores are reported, including whether or not the student data will be part of the CMA for Science reporting and how performance levels are reported for students who used an individualized aid, and how scores are reported under certain conditions (for example, when a student was not tested)
 - General reporting descriptions such as how to calculate number tested
- **Include Indicators**—Defines the appropriate codes to use when a student does not take or complete a test or when a score will not be reported

The scoring specifications are reviewed and revised by the California Department of Education (CDE) and ETS each year. After a version agreeable to all parties is finalized, the CDE issues a formal approval of the scoring and reporting specifications.

Scanning and Scoring

Answer documents are scanned and scored by ETS in accordance with the scoring specifications that have been approved by the CDE. Answer documents are designed to produce a single complete record for each student. This record includes demographic data and scanned responses for each student; once computed, the scored responses and the total test scores for a student are also merged into the same record. All scores, including those available via electronic reporting, must comply with the ETS scoring specifications. ETS has quality control checks in place to ensure the quality and accuracy of scanning and the transfer of scores into the database of student records.

Each local educational agency (LEA) must return scorable and nonscorable materials within five working days after the selected last day of testing for each test administration period.

Types of Scores and Subscores

Raw Score

For all of the tests, the total test raw score equals the number of multiple-choice test items correctly answered.

Subscore

The items in each STS for RLA are aggregated into groups of related content standards to form reporting clusters. A subscore is a measure of an examinee's performance on the items in each reporting cluster. These results are provided only in this technical report. A description of the STS for RLA reporting clusters is provided in Appendix 2.B of Chapter 2, starting on page 19.

Scale Score

Raw scores obtained on each STS for RLA are transformed to three-digit scale scores using the equating process described in Chapter 2 on page 14. Scale scores range from 150 to 600 on each STS for RLA. The scale scores of examinees that have been tested in different years at a given grade level and content area can be compared. However, the raw scores of these examinees cannot be meaningfully compared, because these scores are affected by the relative difficulty of the test taken as well as the ability of the examinee.

Performance Levels

The performance of each student on each STS for RLA is categorized into one of the following performance levels:

- far below basic
- below basic
- basic
- proficient
- advanced

For all STS for RLA, the cut score for the basic performance level is 300 for every test; this means that a student must earn a score of 300 or higher to achieve a basic classification. The cut score for the proficient performance level is 350; this means that a student must earn a score of 350 or higher to achieve a proficient classification. The cut scores for the other performance levels usually vary by grade and content area.

Score Verification Procedures

Various necessary measures are taken to ascertain that the scoring keys are applied to the student responses as intended and that the student scores are computed accurately. In 2014–15, every regular and special-version multiple-choice test is certified by ETS prior to being included in electronic reporting. To certify a test, psychometricians gather a certain number of test cases and verify the accurate application of scoring keys and scoring tables.

Scoring Key Verification Process

Scoring keys, provided in the form planners, are produced by ETS and verified by performing multiple quality-control checks. The form planners contain the information about an assembled test form, including scoring keys, test name, administration year, subscore

identification, and the standards and statistics associated with each item. The quality control checks that are performed before keys are finalized are listed below:

1. Keys in the form planners are checked against their matching test booklets to ensure that the correct keys are listed.
2. The form planners are checked for accuracy against the Form Planner Specification document and the Score Key and Score Conversion document before the keys are loaded into the score key management (SKM) system at ETS.
3. The printed lists of the scoring keys are checked again once the keys have been loaded into the SKM system.
4. The demarcations of various sections in the actual test booklets are checked against the list of demarcations provided by ETS test development staff.
5. Scoring is verified internally at ETS, which generates scores and verifies the scoring of the data by comparing the two results. Any discrepancies are then resolved.
6. The entire scoring system is tested using a test deck that includes typical and extremely atypical response vectors.
7. Classical item analyses are computed on an early sample of data to provide an additional check of the keys. Although rare, if an item is found to be problematic, a follow-up process is carried out for it to be excluded from further analyses.

Overview of Score Aggregation Procedures

In order to provide meaningful results to the stakeholders, STS for RLA scores for a given grade and content area are aggregated at the school, independently testing charter school, district, county, and state levels. The aggregated scores are generated both for individual scores and group scores. The next section contains a description of types of aggregation performed on STS for RLA scores.

Individual Scores

The tables in this section provide state-level summary statistics describing student performance on each STS for RLA.

Score Distributions and Summary Statistics

Summary statistics that describe the performance of students on each STS for RLA in the overall population are presented in Table 7.1.

Included in the tables are the number of items in each test, the number of examinees taking each test, the means and standard deviations of student raw scores, and the means and standard deviations of scale scores. The last two columns in the tables list the raw score means and standard deviations as percentages of the total raw score points in each test.

Table 7.1 Mean and Standard Deviation of Raw and Scale Scores for the STS for RLA (Overall Population)

STS	No. of Items	No. of Examinees	Scale Score		Raw Score		Raw Score %	
			Mean	Std. Dev.	Mean	Std. Dev.	Correct Mean	Correct Std. Dev.
2 RLA	65	1,879	313	49	36.68	12.77	56.44	19.65
3 RLA	65	1,643	314	47	33.81	11.39	52.01	17.52
4 RLA	75	934	310	51	39.02	14.29	52.03	19.05
5 RLA	75	730	313	61	37.06	14.02	49.42	18.70
6 RLA	75	440	311	59	35.79	13.26	47.72	17.67
7 RLA	75	455	318	60	39.50	14.37	52.67	19.17
8 RLA	75	457	317	53	37.32	13.05	49.75	17.40
9 RLA	75	810	315	50	39.73	12.07	52.97	16.09
10 RLA	75	285	316	58	40.32	13.10	53.75	17.47
11 RLA	75	131	311	65	40.92	13.38	54.55	17.85

Note: The overall population consists of the target population plus the optional population.

The percentages of students in each performance level are presented in Table 7.2. The last column of the table presents the overall percentages of examinees that were classified at the proficient level or higher.

The numbers in the summary tables may not match exactly the results reported on the CDE’s Web site because of slight differences in the samples used to compute the statistics. The P2 data file was used for the analyses in this chapter. This file contained the entire test-taking population and all the student records used as of October 28, 2015. This file contained data collected from all LEAs but did not include corrections of demographic data through the California Longitudinal Pupil Achievement Data System. In addition, students with invalid scores were excluded from the tables.

Table 7.2 Percentages of Examinees in Each Performance Level

STS	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Proficient/Advanced *
2 RLA	6%	36%	34%	15%	9%	24%
3 RLA	7%	36%	36%	16%	6%	22%
4 RLA	16%	28%	32%	17%	7%	24%
5 RLA	27%	17%	29%	16%	10%	26%
6 RLA	21%	26%	26%	19%	8%	27%
7 RLA	15%	28%	27%	19%	11%	30%
8 RLA	10%	31%	33%	18%	9%	26%
9 RLA	8%	32%	36%	17%	7%	24%
10 RLA	10%	30%	33%	19%	8%	27%
11 RLA	13%	31%	24%	21%	11%	32%

* May not exactly match the sum of percent proficient and percent advanced due to rounding.

Table 7.A.1 in Appendix 7.A, starting on page 89, show the distributions of scale scores for each STS for RLA for the overall population.

The results are reported in terms of 15 score intervals, each of which contains 30 scale score points. A cell value of “N/A” indicates that there are no obtainable scale scores within that scale-score range for the particular STS for RLA. The distributions are presented for all examinees for each grade level.

Group Scores

Statistics summarizing student performance for each grade-level test for selected groups of students are provided starting on page 90 in Table 7.B.1 through Table 7.B.10 for the STS for RLA. The summary tables are provided for each STS for RLA based on the overall population. When a test is administered at more than one grade level, the results are reported for all students tested and also by grade.

In these tables, students are grouped by demographic characteristics, including gender, economic status, length of enrollment in U.S. schools, and special education programs. The tables show, for each demographic group, the numbers of valid cases, scale score means and standard deviations, the percentages of students in each performance level, as well as the mean-percent correct in each reporting cluster.

Table 7.3 provides definitions of the demographic groups included in the tables. To protect privacy when the number of students in a subgroup is 10 or fewer, the summary statistics at the test- and reporting-cluster-level are not reported and are presented as hyphens. Percentages in these tables may not sum up to 100 due to rounding.

Table 7.3 Subgroup Definitions

Subgroup	Definition
Gender	<ul style="list-style-type: none"> • Male • Female
Economic Status	<ul style="list-style-type: none"> • Not economically disadvantaged • Economically disadvantaged
Enrollment in U.S. Schools	<ul style="list-style-type: none"> • Less than 12 months • 12 months or more
Special Services	<ul style="list-style-type: none"> • No special services • Special services

Reports Produced and Scores for Each Report

The tests that make up the California Assessment of Student Performance and Progress (CAASPP) System provide results or score summaries that are reported for different purposes. The three major purposes are:

1. Communicating with parents and guardians;
2. Informing decisions needed to support student achievement; and
3. Evaluating school programs.

A detailed description of the uses and applications of CAASPP reports is presented in the next section.

Types of Score Reports

There are three categories of STS for RLA reports. These categories and the specific reports in each category are given in Table 7.4.

Table 7.4 Types of STS for RLA Reports

1. Electronic Summary Report	▪ CAASPP Aggregate Report (includes subgroups)
2. Individual Reports	▪ CAASPP Student Data File ▪ CAASPP Student Score Report for STS
3. Internet Reports	▪ STS for RLA Summary Scores (state, county, LEA, school)

The CAASPP aggregate reports and student data files for the LEA are available for the LEA CAASPP coordinator to download from the Test Operations Management System. The LEA forwards the appropriate reports to test sites or, in the case of the CAASPP Student Score Report, sends the report(s) to the child’s parent or guardian and forwards a copy to the student’s school or test site. Reports such as the CAASPP Student Score Reports that include individual student results are not distributed beyond the student’s school. Internet reports are described on the CDE Web site and are accessible to the public online at <http://caaspp.cde.ca.gov/>.

Because results were pre-equated, individual student scores were also available to LEAs prior to the release of final reports via electronic reporting, accessed using the Online Reporting System. This application permits LEAs to view preliminary results data for all tests taken.

Student Score Report Contents

The CAASPP Student Score Report provides scale scores and performance level results for the STS for RLA taken. Scale scores are reported on a scale ranging from 150 to 600. The performance levels reported are: far below basic, below basic, basic, proficient, and advanced.

Reports for students with disabilities who use universal tools, designated supports, or accommodations include a notation that indicates that the student used non-embedded supports (accommodations) or was tested with non-embedded accessibility supports (modifications).

Scores for students who use non-embedded supports are reported in the same way as they are for nonaccommodated students. Non-embedded accessibility supports (modifications), however, change what is being tested and, therefore, change scores. If students use non-embedded accessibility supports (modifications), their scores are counted differently from nonmodified test scores on the CAASPP summary reports—STS for RLA scores for these students are counted as far below basic for tests that report performance levels, regardless of the scale score obtained.

Further information about the CAASPP Student Score Report is provided in Appendix 7.C on page 100.

Student Score Report Applications

STS for RLA results provide parents and guardians with information about their child’s progress. The results are a tool for increasing communication and collaboration between parents or guardians and teachers. Along with report cards from teachers and information from school and classroom tests, the CAASPP Student Score Report can be used by parents and guardians while talking with teachers about ways to improve their child’s achievement of the California content standards.

Schools may use the STS for RLA results to help make decisions about how best to support student achievement. STS for RLA results, however, should never be used as the only source of information to make important decisions about a child's education.

STS for RLA results help LEAs and schools identify strengths and weaknesses in their instructional programs. Each year, LEAs and school staffs examine STS for RLA results for each test administered. Their findings are used to help determine:

- The extent to which students are learning the academic standards,
- Instructional areas that can be improved,
- Teaching strategies that can be developed to address needs of students, and
- Decisions about how to use funds to ensure that students achieve the standards.

Criteria for Interpreting Test Scores

A LEA may use STS for RLA results to help make decisions about student placement, promotion, retention, or other considerations related to student achievement. However, it is important to remember that a single test can provide only limited information. Other relevant information should be considered as well. It is advisable for parents to evaluate their child's strengths and weaknesses in the relevant topics by reviewing classroom work and progress reports in addition to the child's STS for RLA results (CDE, 2015). It is also important to note that a student's score in a content area contains measurement error and could vary somewhat if the student were retested.

Criteria for Interpreting Score Reports

The information presented in various reports must be interpreted with caution when making performance comparisons. When comparing scale score and performance-level results for the STS for RLA, the user is limited to comparisons within the same content area and grade. This is because the score scales are different for each content area and grade. The user may compare scale scores for the same content area and grade, within a school, between schools, or between a school and its district, its county, or the state. The user can also make comparisons within the same grade and content area across years. Comparing scores obtained in different grades or content areas should be avoided because the results are not on the same scale. Comparisons between raw scores or cluster scores should be limited to comparisons within not only content area and grade but also test year. For more details on the criteria for interpreting information provided on the score reports, see the *2015 CAASPP Post-Test Guide* (CDE, 2015).

References

California Department of Education. (2015). *CAASPP 2015 post-test guide*. Sacramento, CA. Retrieved from http://www.caaspp.org/rsc/pdfs/CAASPP.post-test_guide.2015.pdf

Educational Testing Service. (2002). *ETS standards for quality and fairness*. Princeton, NJ: Author.

Appendix 7.A—Scale Score Distribution Tables

In Appendix 7.A, a cell value of “N/A” indicates that there are no obtainable scale scores within that scale-score range for the particular STS for RLA.

In 2014–15, administration of the STS for RLA was voluntary. STS for RLA populations are as follows:

- Overall population—target population plus the optional population
- Target population—students receiving instruction in Spanish or students who have attended school in the United States for less than 12 months (cumulative, not consecutive, months)
- Optional population—students who receive instruction in English and who have attended school in the United States for 12 cumulative months or longer
- In 2014–15, no “English learner (EL) program participation” category information was collected in the individual student level data. Distributions of STS for RLA scale scores are only reported for the overall population.

Table 7.A.1 Distribution of STS for RLA Scale Scores, Grades Two through Eleven (Overall Population)

Scale Score	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11
570 – 600	0	0	0	0	0	0	0	0	0	0
540 – 569	N/A	N/A	0	0	0	0	0	0	0	0
510 – 539	N/A	0	N/A	0	0	0	0	0	0	0
480 – 509	1	0	0	3	2	2	0	0	0	0
450 – 479	8	8	3	9	1	8	3	2	4	0
420 – 449	25	37	12	19	13	20	14	8	12	7
390 – 419	103	55	49	69	32	31	36	58	14	11
360 – 389	224	196	117	66	53	43	52	79	32	16
330 – 359	323	258	157	110	56	76	74	183	52	20
300 – 329	405	394	187	129	74	79	92	161	57	19
270 – 299	392	401	184	126	88	87	87	167	47	17
240 – 269	286	240	164	131	77	79	74	110	39	20
210 – 239	104	50	55	51	37	25	23	29	22	16
180 – 209	8	4	6	17	5	4	2	12	6	4
150 – 179	0	0	0	0	2	1	0	1	0	1

Appendix 7.B—Demographic Summaries

To protect privacy when the number of students in a subgroup is 10 or fewer, the summary statistics at the test- and reporting-cluster-level are not reported and are presented as hyphens in the tables in Appendix 7.B. Percentages in these tables may not sum up to 100 due to rounding.

In 2014–15, administration of the STS for RLA was voluntary. STS populations are as follows:

- Overall population—Target population plus the optional population
- Target population—Students receiving instruction in Spanish or students who have attended school in the United States for less than 12 months (cumulative, not consecutive, months)
- Optional population—Students who receive instruction in English and who have attended school in the United States for 12 cumulative months or longer
- During the 2014–15 administration, no “EL program participation” category information was collected in the individual student level data. Demographic summary is only reported for overall population.

Table 7.B.1 Demographic Summary for RLA, Grade Two (Overall Population)

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level Far Below Basic	Percent in Performance Level Below Basic	Percent in Performance Level Basic	Percent in Performance Level Proficient	Percent in Performance Level Advanced	Mean Percent Correct in Word Analysis and Vocabulary Development	Mean Percent Correct in Reading Comprehension	Mean Percent Correct in Literary Response and Analysis	Mean Percent Correct in Written Conventions	Mean Percent Correct in Writing Strategies
All population valid scores	1,879	313	49	6%	36%	34%	15%	9%	63%	55%	60%	56%	40%
Male	946	308	49	6%	41%	32%	13%	8%	61%	51%	58%	53%	39%
Female	933	319	49	5%	32%	36%	17%	10%	65%	59%	62%	58%	41%
Gender unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
Not economically disadvantaged	158	301	48	9%	47%	29%	9%	6%	59%	50%	58%	48%	36%
Economically disadvantaged	1,721	315	49	6%	35%	34%	16%	9%	63%	55%	60%	56%	41%
Economic status unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
In U.S. schools < 12 months	250	291	47	15%	46%	27%	10%	2%	54%	47%	53%	43%	34%
In U.S. schools ≥ 12 months	1,629	317	49	5%	35%	35%	16%	10%	64%	56%	61%	57%	41%
No special education	1,794	315	49	5%	35%	35%	15%	9%	64%	55%	60%	56%	40%
Special education	85	286	49	16%	51%	19%	11%	4%	50%	46%	48%	44%	32%
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

Table 7.B.2 Demographic Summary for RLA, Grade Three (Overall Population)

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level Far Below Basic	Percent in Performance Level Below Basic	Percent in Performance Level Basic	Percent in Performance Level Proficient	Percent in Performance Level Advanced	Mean Percent Correct in Word Analysis and Vocabulary Development	Mean Percent Correct in Reading Comprehension	Mean Percent Correct in Literary Response and Analysis	Mean Percent Correct in Written Conventions	Mean Percent Correct in Writing Strategies
All population valid scores	1,643	314	47	7%	36%	36%	16%	6%	60%	47%	46%	51%	47%
Male	815	308	44	7%	42%	33%	14%	4%	59%	44%	45%	49%	46%
Female	828	320	48	6%	30%	38%	18%	8%	62%	50%	48%	54%	49%
Gender unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
Not economically disadvantaged	164	322	53	7%	32%	30%	20%	10%	63%	52%	49%	52%	50%
Economically disadvantaged	1,479	313	46	6%	36%	36%	15%	6%	60%	46%	46%	51%	47%
Economic status unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
In U.S. schools < 12 months	306	309	50	10%	37%	31%	16%	6%	60%	49%	46%	44%	44%
In U.S. schools ≥ 12 months	1,337	315	46	6%	36%	37%	16%	6%	61%	47%	47%	53%	48%
No special education	1,563	316	47	6%	35%	36%	16%	6%	61%	48%	47%	52%	48%
Special education	80	288	41	15%	54%	24%	6%	1%	49%	36%	41%	43%	32%
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

Table 7.B.3 Demographic Summary for RLA, Grade Four (Overall Population)

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level Far Below Basic	Percent in Performance Level Below Basic	Percent in Performance Level Basic	Percent in Performance Level Proficient	Percent in Performance Level Advanced	Mean Percent Correct in Word Analysis and Vocabulary Development	Mean Percent Correct in Reading Comprehension	Mean Percent Correct in Literary Response and Analysis	Mean Percent Correct in Written Conventions	Mean Percent Correct in Writing Strategies
All population valid scores	934	310	51	16%	28%	32%	17%	7%	57%	51%	46%	58%	44%
Male	473	301	51	23%	29%	29%	13%	6%	53%	47%	41%	53%	41%
Female	461	320	48	8%	27%	35%	22%	8%	61%	56%	50%	62%	46%
Gender unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
Not economically disadvantaged	142	314	52	14%	29%	30%	18%	9%	58%	53%	48%	58%	46%
Economically disadvantaged	792	310	50	16%	28%	32%	17%	6%	57%	51%	45%	58%	44%
Economic status unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
In U.S. schools < 12 months	319	306	49	16%	30%	32%	16%	6%	54%	51%	46%	54%	44%
In U.S. schools ≥ 12 months	615	312	51	15%	27%	32%	18%	7%	58%	52%	45%	59%	44%
No special education	890	313	50	14%	28%	33%	18%	7%	58%	52%	46%	59%	45%
Special education	44	268	41	45%	36%	16%	2%	0%	43%	35%	34%	38%	27%
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

Table 7.B.4 Demographic Summary for RLA, Grade Five (Overall Population)

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level Far Below Basic	Percent in Performance Level Below Basic	Percent in Performance Level Basic	Percent in Performance Level Proficient	Percent in Performance Level Advanced	Mean Percent Correct in Word Analysis and Vocabulary Development	Mean Percent Correct in Reading Comprehension	Mean Percent Correct in Literary Response and Analysis	Mean Percent Correct in Written Conventions	Mean Percent Correct in Writing Strategies
All population valid scores	730	313	61	27%	17%	29%	16%	10%	50%	42%	53%	54%	48%
Male	390	303	62	35%	18%	26%	11%	10%	48%	40%	50%	50%	45%
Female	340	324	58	19%	16%	33%	21%	11%	54%	45%	57%	58%	51%
Gender unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
Not economically disadvantaged	123	319	66	27%	16%	28%	13%	16%	50%	44%	56%	55%	51%
Economically disadvantaged	607	311	60	27%	17%	30%	16%	9%	50%	42%	53%	54%	47%
Economic status unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
In U.S. schools < 12 months	270	311	64	29%	17%	29%	13%	12%	50%	41%	55%	51%	48%
In U.S. schools ≥ 12 months	460	314	59	26%	18%	30%	17%	9%	50%	43%	52%	55%	48%
No special education	698	315	60	25%	17%	30%	16%	11%	51%	43%	54%	55%	49%
Special education	32	259	46	69%	19%	9%	0%	3%	30%	29%	32%	39%	32%
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

Table 7.B.5 Demographic Summary for RLA, Grade Six (Overall Population)

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level Far Below Basic	Percent in Performance Level Below Basic	Percent in Performance Level Basic	Percent in Performance Level Proficient	Percent in Performance Level Advanced	Mean Percent Correct in Word Analysis and Vocabulary Development	Mean Percent Correct in Reading Comprehension	Mean Percent Correct in Literary Response and Analysis	Mean Percent Correct in Written Conventions	Mean Percent Correct in Writing Strategies
All population valid scores	440	311	59	21%	26%	26%	19%	8%	44%	50%	54%	50%	41%
Male	240	300	56	28%	27%	24%	17%	5%	42%	47%	50%	47%	37%
Female	200	324	59	13%	26%	28%	22%	11%	48%	54%	58%	54%	46%
Gender unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
Not economically disadvantaged	105	314	61	25%	21%	24%	21%	10%	44%	50%	58%	51%	41%
Economically disadvantaged	335	310	58	20%	28%	27%	19%	7%	45%	50%	53%	50%	41%
Economic status unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
In U.S. schools < 12 months	273	314	59	18%	27%	27%	18%	9%	45%	51%	56%	52%	42%
In U.S. schools ≥ 12 months	167	305	58	26%	25%	24%	20%	5%	44%	49%	51%	48%	40%
No special education	435	311	59	21%	26%	26%	19%	8%	45%	50%	54%	50%	41%
Special education	5	–	–	–	–	–	–	–	–	–	–	–	–
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

Table 7.B.6 Demographic Summary for RLA, Grade Seven (Overall Population)

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level Far Below Basic	Percent in Performance Level Below Basic	Percent in Performance Level Basic	Percent in Performance Level Proficient	Percent in Performance Level Advanced	Mean Percent Correct in Word Analysis and Vocabulary Development	Mean Percent Correct in Reading Comprehension	Mean Percent Correct in Literary Response and Analysis	Mean Percent Correct in Written Conventions	Mean Percent Correct in Writing Strategies
All population valid scores	455	318	60	15%	28%	27%	19%	11%	58%	51%	51%	54%	51%
Male	221	307	59	21%	30%	26%	14%	9%	55%	48%	48%	49%	46%
Female	234	328	58	9%	27%	27%	23%	14%	61%	54%	53%	59%	55%
Gender unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
Not economically disadvantaged	98	320	60	9%	34%	26%	17%	14%	58%	51%	53%	56%	52%
Economically disadvantaged	357	317	60	16%	27%	27%	19%	10%	58%	51%	50%	54%	50%
Economic status unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
In U.S. schools < 12 months	323	322	58	11%	27%	31%	18%	12%	60%	53%	53%	55%	51%
In U.S. schools ≥ 12 months	132	308	62	23%	31%	17%	20%	9%	54%	48%	46%	51%	49%
No special education	450	319	59	14%	29%	27%	19%	11%	58%	51%	51%	55%	51%
Special education	5	–	–	–	–	–	–	–	–	–	–	–	–
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

Table 7.B.7 Demographic Summary for RLA, Grade Eight (Overall Population)

	Number Tested	Mean Scale Score	Std. Dev. of Number Correct Scores	Percent in Performance Level Far Below Basic	Percent in Performance Level Below Basic	Percent in Performance Level Basic	Percent in Performance Level Proficient	Percent in Performance Level Advanced	Mean Percent Correct in Word Analysis and Vocabulary Development	Mean Percent Correct in Reading Comprehension	Mean Percent Correct in Literary Response and Analysis	Mean Percent Correct in Written Conventions	Mean Percent Correct in Writing Strategies
All population valid scores	457	317	53	10%	31%	33%	18%	9%	54%	45%	50%	59%	43%
Male	243	311	55	13%	35%	28%	17%	7%	53%	44%	49%	56%	41%
Female	214	323	51	7%	27%	39%	18%	10%	54%	46%	50%	63%	46%
Gender unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
Not economically disadvantaged	113	327	56	8%	27%	34%	18%	14%	59%	50%	53%	61%	46%
Economically disadvantaged	344	314	52	10%	32%	33%	17%	7%	52%	44%	49%	59%	43%
Economic status unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
In U.S. schools < 12 months	333	321	54	9%	28%	35%	17%	11%	56%	47%	51%	61%	44%
In U.S. schools ≥ 12 months	124	306	50	11%	39%	27%	19%	3%	49%	42%	46%	56%	41%
No special education	449	318	53	10%	30%	34%	18%	9%	54%	46%	50%	60%	44%
Special education	8	–	–	–	–	–	–	–	–	–	–	–	–
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

Table 7.B.8 Demographic Summary for RLA, Grade Nine (Overall Population)

	Number Tested	Mean Scale Score	Std. Dev. of Number Correct Scores	Percent in Performance Level Far Below Basic	Percent in Performance Level Below Basic	Percent in Performance Level Basic	Percent in Performance Level Proficient	Percent in Performance Level Advanced	Mean Percent Correct in Word Analysis and Vocabulary Development	Mean Percent Correct in Reading Comprehension	Mean Percent Correct in Literary Response and Analysis	Mean Percent Correct in Written Conventions	Mean Percent Correct in Writing Strategies
All population valid scores	810	315	50	8%	32%	36%	17%	7%	71%	55%	55%	51%	44%
Male	468	308	50	10%	34%	35%	14%	6%	69%	53%	53%	48%	42%
Female	342	325	48	4%	28%	38%	21%	9%	73%	58%	58%	55%	47%
Gender unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
Not economically disadvantaged	273	315	51	8%	30%	37%	17%	7%	70%	55%	56%	51%	44%
Economically disadvantaged	537	315	49	7%	33%	36%	17%	7%	71%	55%	54%	51%	44%
Economic status unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
In U.S. schools < 12 months	703	317	50	7%	31%	36%	18%	8%	71%	56%	55%	51%	45%
In U.S. schools ≥ 12 months	107	302	48	10%	36%	38%	10%	5%	66%	50%	51%	49%	40%
No special education	809	315	50	7%	32%	36%	17%	7%	71%	55%	55%	51%	44%
Special education	1	–	–	–	–	–	–	–	–	–	–	–	–
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

Table 7.B.9 Demographic Summary for RLA, Grade Ten (Overall Population)

	Number Tested	Mean Scale Score	Std. Dev. of Number Correct Scores	Percent in Performance Level Far Below Basic	Percent in Performance Level Below Basic	Percent in Performance Level Basic	Percent in Performance Level Proficient	Percent in Performance Level Advanced	Mean Percent Correct in Word Analysis and Vocabulary Development	Mean Percent Correct in Reading Comprehension	Mean Percent Correct in Literary Response and Analysis	Mean Percent Correct in Written Conventions	Mean Percent Correct in Writing Strategies
All population valid scores	285	316	58	10%	30%	33%	19%	8%	66%	58%	56%	53%	44%
Male	168	308	57	11%	35%	32%	15%	6%	63%	56%	54%	50%	41%
Female	117	327	59	8%	23%	35%	24%	10%	69%	61%	59%	58%	47%
Gender unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
Not economically disadvantaged	61	328	57	5%	31%	33%	20%	11%	72%	62%	61%	55%	47%
Economically disadvantaged	224	312	58	11%	30%	33%	19%	7%	64%	57%	55%	53%	43%
Economic status unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
In U.S. schools < 12 months	194	323	57	8%	25%	37%	21%	9%	69%	61%	59%	56%	44%
In U.S. schools ≥ 12 months	91	300	58	14%	41%	25%	14%	5%	59%	52%	50%	48%	42%
No special education	284	316	58	10%	30%	33%	19%	8%	66%	58%	56%	53%	44%
Special education	1	–	–	–	–	–	–	–	–	–	–	–	–
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

Table 7.B.10 Demographic Summary for RLA, Grade Eleven (Overall Population)

	Number Tested	Mean Scale Score	Std. Dev. of Number Correct Scores	Percent in Performance Level Far Below Basic	Percent in Performance Level Below Basic	Percent in Performance Level Basic	Percent in Performance Level Proficient	Percent in Performance Level Advanced	Mean Percent Correct in Word Analysis and Vocabulary Development	Mean Percent Correct in Reading Comprehension	Mean Percent Correct in Literary Response and Analysis	Mean Percent Correct in Written Conventions	Mean Percent Correct in Writing Strategies
All population valid scores	131	311	65	13%	31%	24%	21%	11%	59%	49%	58%	63%	52%
Male	65	289	64	22%	38%	20%	11%	9%	53%	46%	50%	55%	45%
Female	66	333	59	5%	24%	27%	30%	14%	65%	52%	65%	71%	59%
Gender unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
Not economically disadvantaged	22	310	74	23%	23%	23%	18%	14%	63%	48%	60%	58%	50%
Economically disadvantaged	109	311	64	11%	33%	24%	21%	11%	58%	49%	57%	64%	53%
Economic status unknown	0	–	–	–	–	–	–	–	–	–	–	–	–
In U.S. schools < 12 months	97	318	63	7%	33%	26%	23%	11%	61%	50%	59%	65%	55%
In U.S. schools ≥ 12 months	34	293	68	29%	26%	18%	15%	12%	54%	45%	55%	57%	44%
No special education	131	311	65	13%	31%	24%	21%	11%	59%	49%	58%	63%	52%
Special education	0	–	–	–	–	–	–	–	–	–	–	–	–
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

Appendix 7.C—Types of Score Reports

Table 7.C.1 Score Reports Reflecting STS for RLA Results

2014–15 CAASPP Student Score Reports	
Description	Use and Distribution
<p>The CAASPP Student Score Report—STS for RLA A report in Spanish for the STS for RLA for the student’s grade level (grades two through eleven)</p>	
<p>This report provides parents/guardians and teachers with the student’s results, presented in tables and graphs. Data presented include the following:</p> <ul style="list-style-type: none"> • Scale scores • Performance levels 	<p>This report includes individual student results and is not distributed beyond parents/guardians and the student’s school.</p> <p>Two copies of this report are provided for each student. One is for the student’s current teacher and one is to be distributed by the LEA to parents/guardians.</p>
<p>Subgroup Summary</p>	
<p>This set of reports disaggregates and reports results by the following subgroups:</p> <ul style="list-style-type: none"> • All students • Disability status • Economic status • Gender • English proficiency • Primary ethnicity • Economic status <p>These reports contain no individual student-identifying information and are aggregated at the school, LEA, county, and state levels.</p> <p>For each subgroup within a report and for the total number of students, the following data are included for each test:</p> <ul style="list-style-type: none"> • Total number tested in the subgroup • Percent of enrollment tested in the subgroup • Number and percent of valid scores • Number tested who received scores • Mean scale score • Standard deviation of scale score • Number and percent of students scoring at each performance level 	<p>This report is a resource for evaluators, researchers, teachers, parents/guardians, community members, and administrators.</p> <p>Each LEA can download this report for the whole LEA and the schools within it from the Test Operations Management System.</p> <p>Note: The data on this report may be shared with parents/guardians, community members, and the media only if the data are for 11 or more students.</p>

Chapter 8: Analyses

This chapter summarizes the item- and test-level statistics obtained for the Standards-based Tests in Spanish (STS) for Reading/Language Arts (RLA) administered during the spring of 2014–15 test administration.

The statistics presented in this chapter are divided into three sections in the following order:

1. Classical Item Analyses
2. Reliability Analyses
3. Item Response Theory (IRT) Analyses

Prior to the 2013–14 administration, differential item functioning (DIF) analyses were performed based on the final item analysis (FIA) sample for all operational and field-test items to assess differences in the item performance of groups of students that differ in their demographic characteristics. In 2014–15 administration, because the intact forms were used, DIF analyses were not performed.

Each of the sets of analyses on data from the 2014–15 administration is presented in the body of the text and in the appendices as listed below.

1. Appendix 8.A on page 119 presents the classical item analyses, including proportion-correct value (p -value) and point-biserial correlation (Pt-Bis) for each item in each operational test. Because all forms were either intact or modified with replacement items, p -values and Pt-Bis are shown for both the original and the current administration of the tests. In addition, the average and median p -value and Pt-Bis for each operational test based on their current administration are presented in Table 8.1 on page 102.
2. Appendix 8.B on page 125 presents results of the reliability analyses of total test scores and subscores for the target population as a whole and for selected subgroups within the target population. Also presented are results of the analyses of the accuracy and consistency of the performance classifications.
3. Appendix 8.C on page 137 presents the summaries of Rasch item difficulty statistics (b -values) for the operational items for the forms for the STS for RLA in grades six and eight, which contain replacement items. (For the summaries of b -values, refer to Appendix D of the *STS Technical Report* for the year each grade-level RLA form was administered originally.) In addition, the appendix presents the scoring tables obtained as a result of the IRT equating process.

Samples Used for the Analyses

STS for RLA analyses were conducted at different times after test administration and involved varying proportions of the full STS for RLA data. The majority of the analyses presented in this chapter, including the classical item analyses presented in Appendix 8.A and the reliability statistics included in Appendix 8.B were calculated on the STS overall population using the P2 data file, which contained the entire test-taking population and all the student records used as of October 28, 2015. This file contained data collected from all local educational agencies (LEAs) but did not include corrections of demographic data through the California Longitudinal Pupil Achievement Data System. In addition, students with invalid scores were excluded.

During the 2014–15 administration, neither IRT calibrations nor scaling are implemented because intact forms were reused and results were pre-equated. For the reused intact

forms, the IRT results were derived based on the equating sample of the previous administration which can be found in Appendix D of the *STS Technical Report* in the year the form was administered originally; see Table 8.4 on page 111 for administration years.

Classical Item Analyses

Multiple-Choice Items

The classical item statistics that included overall and item-by-item proportion-correct indices and the point-biserial correlation indices were computed for the operational items. The p -value of an item represents the proportion of examinees in the sample that answered an item correctly. The formula for p -value is:

$$p\text{-value}_i = \frac{N_{ic}}{N_i} \quad (8.1)$$

where,

N_{ic} is the number of examinees that answered item i correctly, and

N_i is the total number of examinees that attempted the item.

The point-biserial correlation is a special case of the Pearson product-moment correlation used to measure the strength of the relationship between two variables, one dichotomously and one continuously measured—in this case, the item score (right/wrong) and the total test score. The formula for the Pearson product-moment correlation is:

$$r_{X_i T} = \frac{\text{cov}(X_i, T)}{s_{X_i} s_T} \quad (8.2)$$

where,

$\text{cov}(X_i, T)$ is the covariance between the score of item i and total score T ,

s_{X_i} is the standard deviation for the score of item i , and

s_T is the standard deviation for total score T .

The classical statistics for the current administration of the overall test are presented in Table 8.1. The item-by-item values for the classical statistics, including p -values, point-biserial correlations, distributional percents, and mean scores, are presented in Table 8.A.1 through Table 8.A.4, starting on page 119. Each set of values is presented for both the current and the original presentation of each STS for RLA.

Table 8.1 Mean and Median Proportion Correct and Point-Biserial Correlation

STS	No. of Items	No. of Examinees	Mean		Median	
			p -value	Pt-Bis	p -value	Pt-Bis
2 RLA	65	1,879	0.56	0.41	0.55	0.42
3 RLA	65	1,643	0.52	0.37	0.52	0.37
4 RLA	75	934	0.52	0.39	0.52	0.41
5 RLA	75	730	0.49	0.38	0.50	0.38
6 RLA	75	440	0.48	0.36	0.48	0.35
7 RLA	75	455	0.53	0.40	0.52	0.41
8 RLA	75	457	0.50	0.36	0.48	0.38

STS	No. of Items	No. of Examinees	Mean		Median	
			<i>p</i> -value	Pt-Bis	<i>p</i> -value	Pt-Bis
9 RLA	75	810	0.53	0.34	0.52	0.35
10 RLA	75	285	0.54	0.37	0.54	0.38
11 RLA	75	131	0.55	0.37	0.54	0.38

Reliability Analyses

Reliability focuses on the extent to which differences in test scores reflect true differences in the knowledge, ability, or skill being tested, rather than fluctuations due to chance or random factors. The variance in the distribution of test scores—essentially, the differences among individuals—is partly due to real differences in the knowledge, skill, or ability being tested (true-score variance) and partly due to random unsystematic errors in the measurement process (error variance).

The number used to describe reliability is an estimate of the proportion of the total variance that is true-score variance. Several different ways of estimating this proportion exist. The estimates of reliability reported here are internal-consistency measures, which are derived from analysis of the consistency of the performance of individuals on items within a test (internal-consistency reliability). Therefore, they apply only to the test form being analyzed. They do not take into account form-to-form variation due to equating limitations or lack of parallelism, nor are they responsive to day-to-day variation due, for example, to students' state of health or testing environment.

Reliability coefficients can range from 0 to 1. The higher the reliability coefficient for a set of scores, the more likely individuals would be to obtain very similar scores if they were retested. The formula for the internal-consistency reliability as measured by Cronbach's Alpha (Cronbach, 1951) is defined by equation 8.3:

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n s_i^2}{s_t^2} \right] \quad (8.3)$$

where,

n is the number of items,

s_i^2 is the variance of scores on the item i , and

s_t^2 is the variance of the total score.

The standard error of measurement (SEM) provides a measure of score instability in the score metric. The SEM was computed as shown in equation 8.4:

$$s_e = s_t \sqrt{1 - \alpha} \quad (8.4)$$

where,

α is the reliability estimated in equation 8.3, and

s_t is the standard deviation (SD) of the total score (either the total raw score or scale score).

The SEM is particularly useful in determining the confidence interval (CI) that captures an examinee's true score. Assuming that measurement error is normally distributed, it can be

said that upon infinite replications of the testing occasion, approximately 95 percent of the CIs of ± 1.96 SEM around the observed score would contain an examinee's true score (Crocker & Algina, 1986). For example, if an examinee's observed score on a given test equals 15 points, and the SEM equals 1.92, one can be 95 percent confident that the examinee's true score lies between 11 and 19 points (15 ± 3.76 rounded to the nearest integer).

Table 8.2 gives the reliability and SEM for each of the STS for RLA, along with the number of items and examinees upon which those analyses were performed.

Table 8.2 Reliabilities and SEMs for the STS for RLA

STS	Test Length	No. Examinees	Reliab.	Scale Score			Raw Score		
				Mean	SD	SEM	Mean	SD	SEM
2 RLA	65	1,879	0.92	313	49	13.61	36.68	12.77	3.53
3 RLA	65	1,643	0.90	314	47	14.94	33.81	11.39	3.63
4 RLA	75	934	0.93	310	51	13.72	39.02	14.29	3.87
5 RLA	75	730	0.92	313	61	17.01	37.06	14.02	3.91
6 RLA	75	440	0.91	311	59	17.56	35.79	13.26	3.97
7 RLA	75	455	0.93	318	60	16.12	39.50	14.37	3.88
8 RLA	75	457	0.91	317	53	16.10	37.32	13.05	3.94
9 RLA	75	810	0.90	315	50	15.87	39.73	12.07	3.85
10 RLA	75	285	0.91	316	58	17.32	40.32	13.10	3.89
11 RLA	75	131	0.91	311	65	19.07	40.92	13.38	3.90

Subgroup Reliabilities and SEMs

The reliabilities of the STS for RLA were examined for various subgroups of the examinee population. The subgroups included in these analyses were defined by their gender, economic status, provision of special services and length of attendance in U.S. schools.

Reliabilities and SEM information for the total test scores and the reporting cluster scores are reported for each subgroup analysis. Table 8.B.2 through Table 8.B.9 present the overall test reliabilities for the various subgroups. Table 8.B.10 presents the cluster-level reliabilities for the subgroups based on gender and economic status. The next table, Table 8.B.11, shows the same analyses for the subgroups based on provision of special services and the length of attendance in U.S. schools.

Note that the reliabilities are reported only for samples that comprise 11 or more examinees. Also, in some cases, score reliabilities were not estimable and are presented in the tables as hyphens. Finally, results based on samples that contain 50 or fewer examinees should be interpreted with caution due to small sample sizes.

Conditional Standard Errors of Measurement

As part of the IRT-based equating procedures, scale-score conversion tables and conditional standard errors of measurement (CSEMs) are produced. CSEMs for STS for RLA scale scores are based on IRT and are calculated by the IRTEQUATE module in a computer system called the Generalized Analysis System (GENASYS).

The CSEM is estimated as a function of measured ability. It is typically smaller in scale-score units toward the center of the scale in the test metric, where more items are located, and larger at the extremes, where there are fewer items. An examinee's CSEM under the IRT framework is equal to the inverse of the square root of the test information function:

$$\text{CSEM}(\hat{\theta}) = \frac{1}{\sqrt{I(\hat{\theta})}} a \quad (8.5)$$

where,

$\text{CSEM}(\hat{\theta})$ is the standard error of measurement, and

$I(\hat{\theta})$ is the test information function at ability level $\hat{\theta}$.

The statistic is multiplied by a , where a is the original scaling factor needed to transform theta to the scale-score metric. The value of a varies by grade level.

SEMs vary across the scale. When a test has cut scores, it is important to provide CSEMs at the cut scores.

Table 8.3 presents the scale score CSEMs at the lowest score required for a student to be classified in the below basic, basic, proficient, and advanced performance levels for each STS for RLA.

The CSEMs tend to be higher at the advanced cut points for all tests. The pattern of lower values of CSEMs at the basic and proficient levels are expected since (1) more items tend to be of middle difficulty; and (2) items at the extremes still provide information toward the middle of the scale. This results in more precise scores in the middle of the scale and less precise scores in the extremes of the scale.

Table 8.3 Scale Score CSEM at Performance-level Cut Points

STS	Below Basic		Basic		Proficient		Advanced	
	Min. SS	CSEM	Min. SS	CSEM	Min. SS	CSEM	Min. SS	CSEM
2 RLA	242	14	300	12	350	14	386	17
3 RLA	251	15	300	14	350	15	393	18
4 RLA	256	14	300	13	350	14	387	16
5 RLA	271	17	300	16	350	17	401	19
6 RLA	260	18	300	17	350	17	401	19
7 RLA	256	16	300	15	350	16	399	18
8 RLA	248	17	300	15	350	16	401	18
9 RLA	248	16	300	15	350	16	396	18
10 RLA	240	18	300	16	350	17	394	19
11 RLA	235	20	300	18	350	19	397	21

Decision Classification Analyses

The methodology used for estimating the reliability of classification decisions is described in Livingston and Lewis (1995) and is implemented using the Educational Testing Service (ETS)-proprietary computer program RELCLASS-COMP (Version 4.14).

Decision accuracy describes the extent to which examinees are classified in the same way as they would be on the basis of the average of all possible forms of a test. Decision accuracy answers the following question: How does the actual classification of test-takers, based on their single-form scores, agree with the classification that would be made on the basis of their true scores, if their true scores were somehow known? RELCLASS-COMP estimates decision accuracy using an estimated multivariate distribution of reported classifications on the current form of the exam and the classifications based on an all-forms average (true score).

Decision consistency describes the extent to which examinees are classified in the same way as they would be on the basis of a single form of a test other than the one for which data are available. Decision consistency answers the following question: What is the agreement between the classifications based on two nonoverlapping, equally difficult forms of the test? RELCLASS-COMP also estimates decision consistency using an estimated multivariate distribution of reported classifications on the current form of the test and classifications on a hypothetical alternate form using the reliability of the test and strong true-score theory.

In each case, the proportion of classifications with exact agreement is the sum of the entries in the diagonal of the contingency table representing the multivariate distribution. Reliability of classification at a cut score is estimated by collapsing the multivariate distribution at the passing score boundary into an n by n table (where n is the number of performance levels) and summing the entries in the diagonal. Figure 8.1 and Figure 8.2 present the two scenarios graphically.

Figure 8.1 Decision Accuracy for Achieving a Performance Level

		Decision made on the form taken	
		Does not achieve a performance level	Achieves a performance level
True status on all-forms average	Does not achieve a performance level	Correct classification	Misclassification
	Achieves a performance level	Misclassification	Correct classification

Figure 8.2 Decision Consistency for Achieving a Performance Level

		Decision made on the alternate form taken	
		Does not achieve a performance level	Achieves a performance level
Decision made on the form taken	Does not achieve a performance level	Correct classification	Misclassification
	Achieves a performance level	Misclassification	Correct classification

The results of these analyses are presented in Table 8.B.12 through Table 8.B.21 in Appendix 8.B, starting on page 133.

Each table includes the contingency tables for both accuracy and consistency of the various performance-level classifications. The proportion of students being accurately classified is determined by summing across the diagonals of the upper tables. The proportion of consistently classified students is determined by summing the diagonals of the lower tables.

The classifications are collapsed to below-proficient versus proficient and above.

Validity Evidence

Validity refers to the degree to which each interpretation or use of a test score is supported by evidence that is gathered (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014; ETS, 2002). It is a central concern underlying the development, administration, and scoring of a test and the uses and interpretations of test scores.

Validation is the process of accumulating evidence to support each proposed score interpretation or use. It involves more than a single study or gathering of one particular kind of evidence. Validation involves multiple investigations and various kinds of evidence (AERA, APA, & NCME, 2014; Cronbach, 1971; ETS, 2002; Kane, 2006). The process begins with test design and continues through the entire assessment process, including item development and field testing, analyses of item and test data, test scaling, scoring, and score reporting.

This section presents the evidence gathered to support the intended uses and interpretations of scores for the STS for RLA testing program. The description is organized in the manner prescribed by *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999). These standards require a clear definition of the purpose of the test, which includes a description of the qualities—called constructs—that are to be assessed by a test, the population to be assessed, as well as how the scores are to be interpreted and used.

In addition, the *Standards* identify five kinds of evidence that can provide support for score interpretations and uses, which are as follows:

1. Evidence based on test content;
2. Evidence based on relations to other variables;
3. Evidence based on response processes;
4. Evidence based on internal structure; and
5. Evidence based on the consequences of testing.

These kinds of evidence are also defined as important elements of validity information in documents developed by the U.S. Department of Education (USD OE) for the peer review of testing programs administered by states in response to the Elementary and Secondary Education Act (USD OE, 2001).

The next section defines the purpose of the STS for RLA, followed by a description and discussion of the kinds of validity evidence that have been gathered.

Purpose of the STS for RLA

As mentioned in Chapter 1, the purpose of the STS for RLA program is to permit students to demonstrate achievement of the California content standards in reading/language arts—approved in 1997 by the State Board of Education (SBE)—through a primary language test in Spanish.

The Constructs to Be Measured

The STS for RLA, administered in Spanish, are designed to show how well students perform relative to the California content standards. These content standards were approved in 1997 by the SBE; they describe what students should know and be able to do at each grade level.

Test blueprints and specifications written to define the procedures used to measure the content standards provide an operational definition of the construct to which each set of standards refers—that is, they define, for each content area to be assessed, the tasks to be presented, the administration instructions to be given, and the rules used to score examinee responses. They control as many aspects of the measurement procedure as possible so that the testing conditions will remain the same over test administrations (Cronbach, 1971; Cronbach, Gleser, Nanda, & Rajaratnam, 1972) to minimize construct-irrelevant score

variance (Messick, 1989). The content blueprints for the STS for RLA can be found on the CDE STAR STS Blueprints Web page at <http://www.cde.ca.gov/ta/tg/sr/stsblueprints.asp>. ETS developed all STS for RLA test items to conform to the SBE-approved content standards and test blueprints.

Interpretations and Uses of the Scores Generated

Total test scores expressed as scale scores and student performance levels are generated for each grade-level and content-area test. The total test scale score is used to draw inferences about a student's achievement in the content area and to classify the achievement into one of five performance levels: advanced, proficient, basic, below basic, and far below basic.

Reporting cluster scores, also called subscores, are used to draw inferences about a student's achievement in each of several specific knowledge or skill areas covered by each test. In past years, when cluster results were reported, the results compared an individual student's percent-correct score to the average percent-correct for the state as a whole. The range of scores for students who scored proficient on the total test was also provided for each cluster using a percent-correct metric. The reference points for this range were: (1) the average percent-correct for students who received the lowest score qualifying for the proficient performance level; and (2) the average percent-correct for students who received the lowest score qualifying for the advanced performance level, minus one percent. A detailed description of the uses and applications of STS for RLA scores as used in past years is presented in Chapter 7, which starts on page 81. Note that these were not used in reporting student results to LEAs or test sites, or in Student Score Reports.

The tests that make up the CAASPP System, along with other assessments, provide results or score summaries that are used for different purposes. The three major purposes of the STS for RLA are:

1. Communicating with parents and guardians;
2. Informing decisions needed to support student achievement; and
3. Evaluating school programs.

These are the only uses and interpretations of scores for which validity evidence has been gathered. If the user wishes to interpret or use the scores in other ways, the user is cautioned that the validity of doing so has not been established (AERA, APA, & NCME, 2014, Standard 1.3). The user is advised to gather evidence to support these additional interpretations or uses (AERA, APA, & NCME, 2014, Standard 1.4).

Intended Test Population(s)

The STS for RLA are optional tests that are targeted toward Spanish-speaking English learners (ELs) who have been in school in the United States for less than 12 cumulative months or who receive instruction in Spanish. However, all students who are ELs and whose primary language is Spanish are eligible to take the STS for RLA. The two distinct STS populations are the "target" and "nontarget/optional" populations. The target population consists of students receiving instruction in Spanish or students who have attended school in the United States for less than 12 months. These are cumulative, not necessarily consecutive, months. The nontarget/optional population consists of students who receive instruction in English and who have attended school in the United States for 12 cumulative months or longer.

Students in grades two through eleven are tested in RLA. Only those students whose parents/guardians have submitted written requests to exempt them from CAASPP System testing do not take an STS for RLA.

See the subsection “Intended Population” on page 1 for additional information about the intended test population.

Validity Evidence Collected

Evidence Based on Content

According to *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), analyses that demonstrate a strong relationship between the test content and the construct that the test was designed to measure can provide important evidence of validity. In current K–12 testing, the construct of interest usually is operationally defined by state content standards and the test blueprints that specify the content, format, and scoring of items that are admissible measures of the knowledge and skills described in the content standards. Evidence that the items meet these specifications and represent the domain of knowledge and skills referenced by the standards supports the inference that students’ scores on these items can appropriately be regarded as measures of the intended construct.

As noted in the AERA, APA, and NCME *Standards* (2014), evidence based on test content may involve logical analyses of test content in which experts judge the adequacy with which the test content conforms to the test specifications and represents the intended domain of content. Such reviews can also be used to determine whether the test content contains material that is not relevant to the construct of interest. Analyses of test content may also involve the use of empirical evidence of item quality.

Also to be considered in evaluating test content are the procedures used for test administration and test scoring. As Kane (2006, p. 29) has noted, although evidence that appropriate administration and scoring procedures have been used does not provide compelling evidence to support a particular score interpretation or use, such evidence may prove useful in refuting rival explanations of test results. Evidence based on content includes the following:

Description of the state standards—As was noted in Chapter 1, the SBE adopted rigorous content standards in 1997 and 1998 in four major content areas: English–language arts, history–social science, mathematics, and science. These standards were designed to guide instruction and learning for all students in the state and to bring California students to world-class levels of achievement. The STS program was instituted to permit students to demonstrate achievement of the content standards in reading/language arts and mathematics using a Spanish-language test. The content standards for English–language arts adopted in 1997 guided the development of the STS for RLA.

Specifications and blueprints—ETS maintains item specifications for each STS for RLA. The item specifications describe the characteristics of the items that should be written to measure each content standard. A thorough description of the specifications can be found in Chapter 3, starting on page 37. Once the items are developed and field-tested, ETS selects all STS for RLA test items to conform to the SBE-approved California content standards and test blueprints. Test blueprints for the STS for RLA were proposed by ETS and reviewed and approved by the Assessment Review Panels (ARPs), which are advisory panels to the CDE and ETS on areas related to item development for the STS for

RLA. Test blueprints were also reviewed and approved by the CDE and presented to the SBE for adoption. There have been no recent changes in the blueprints for the STS for RLA. The test blueprints for the STS for RLA can be found on the CDE STAR STS Blueprints Web page at <http://www.cde.ca.gov/ta/tg/sr/stsblueprints.asp>.

Item development process—A detailed description of the item development process for the STS for RLA is presented in Chapter 3, starting on page 37.

Item review process—Chapter 3 explains in detail the extensive item review process applied to items written for use in the STS for RLA. In brief, items written for the STS for RLA underwent multiple review cycles and involved multiple groups of reviewers. One of the reviews was carried out by an external reviewer, that is, the ARPs. The ARPs were responsible for reviewing all newly developed items for alignment to the California content standards.

Form construction process—For each test, the content standards, blueprints, and test specifications were used as the basis for choosing items (refer to “Test Assembly” on page 9 in Chapter 2 for the replacement item selection rules). Additional targets for item difficulty and discrimination that were used for test construction were defined in light of what are desirable statistical characteristics in test items and statistical evaluations of the STS for RLA items.

Guidelines for test construction were established with the goal of maintaining parallel forms to the greatest extent possible from year to year. Details can be found in Chapter 4, starting on page 47.

Additionally, an external review panel, the Statewide Pupil Assessment Review (SPAR), was responsible for reviewing and approving the achievement tests to be used statewide for the testing of students in California public schools, grades two through eleven. More information about the SPAR is given in Chapter 3, starting on page 43.

Evidence Based on Relations to Other Variables

Empirical results concerning the relationships between the score on a test and measures of other variables external to the test can also provide evidence of validity when these relationships are found to be consistent with the definition of the construct that the test is intended to measure. As indicated in the *Standards* (AERA, APA, & NCME, 2014), the variables investigated can include other tests that measure the same construct and different constructs, criterion measures that scores on the test are expected to predict, as well as demographic characteristics of examinees that are expected to be related and unrelated to test performance.

Differential Item Functioning Analyses

Analyses of DIF can provide evidence of the degree to which a score interpretation or use is valid for individuals who differ in particular demographic characteristics. For the STS for RLA, DIF analyses were performed after the test forms’ original administration on all operational items and all field-test items for which sufficient student samples were available.

The results of the DIF analyses are presented in Appendix 8.E of the *STS Technical Report* produced for the year the form was originally administered. Reports are linked on the CDE Technical Reports and Studies Web page at <http://www.cde.ca.gov/ta/tg/sr/technicalrpts.asp>. The year of original administration for each STS for RLA is shown in Table 8.4. Please note that the STS for RLA in grades six and eight used the forms for the

2013–14 administration; these two were modified intact forms containing six replacement items in total from the original forms administered in 2011–12.

Table 8.4 Original Year of Administration for the STS for RLA

STS	Year
2 RLA	2013
3 RLA	2013
4 RLA	2013
5 RLA	2013
6 RLA	2012
7 RLA	2013
8 RLA	2012
9 RLA	2011
10 RLA	2013
11 RLA	2013

Evidence Based on Response Processes

As noted in the APA, AERA, and NCME *Standards* (2014), additional support for a particular score interpretation or use can be provided by theoretical and empirical evidence indicating that examinees are using the intended response processes when responding to the items in a test. This evidence may be gathered from interacting with examinees in order to understand what processes underlie their item responses.

Evidence Based on Internal Structure

As suggested by the *Standards* (AERA, APA, & NCME, 2014), evidence of validity can also be obtained from studies of the properties of the item scores and the relationship between these scores and scores on components of the test. To the extent that the score properties and relationships found are consistent with the definition of the construct measured by the test, support is gained for interpreting these scores as measures of the construct.

For the STS for RLA, it is assumed that a single construct underlies the total scores obtained on each test. Evidence to support this assumption can be gathered from the results of item analyses, evaluations of internal consistency, and studies of reliability.

With respect to the subscores that are reported, these scores are intended to reflect examinees' knowledge and/or skill in an area that is part of the construct underlying the total test. Analyses of the intercorrelations among the subscores themselves and between the subscores and total test score can be used for studying this aspect of the construct. Information about the internal consistency of the items on which each subscore is based is also useful to provide.

Classical Statistics

Point-biserial correlations calculated for the items in a test show the degree to which the items discriminate between students with low and high scores on a test. To the degree that the correlations are high, evidence that the items assess the same construct is provided. As shown in Table 8.1, the mean point biserial was between 0.34 and 0.41. The point biserials for the individual items in the STS for RLA are presented in Table 8.A.1 through Table 8.A.4.

Also germane to the validity of a score interpretation are the ranges of item difficulty for the items on which a test score will be based. The finding that items have difficulties that span

the range of examinee ability provides evidence that examinees at all levels of ability are adequately measured by the items. Information on average item p -values is given in Table 8.1; individual item p -values are presented in Table 8.A.1 through Table 8.A.4 side by side with the p -values of these items obtained when the intact or modified intact forms were originally used. (The summaries of the IRT difficulty indicator b -values for operational items can be found in Appendix D of the *STS Technical Report* for the year in which each grade-level RLA form was administered originally; see Table 8.4 on page 111 for administration years.)

The data in Table 8.1 indicate that STS for RLA tests had average p -values that range from 0.48 to 0.56.

Reliability

Reliability is a prerequisite for validity. The finding of reliability in student scores supports the validity of the inference that the scores reflect a stable construct. This section will describe briefly findings concerning the total test level, as well as reliability results for the reporting clusters.

Overall reliability—The reliability analyses on each of the operational STS for RLA are presented in Table 8.2. The results indicate that the reliabilities for the STS for RLA in grades two through eleven were relatively high, ranging from 0.90 to 0.93.

Reporting cluster reliabilities—For each STS for RLA, number-correct scores are computed for the reporting clusters. The reliabilities of these scores are presented in Table 8.B.1. The reliabilities of reporting clusters are invariably lower than those for the total tests since they are based on fewer items. Consistent with the findings of previous years, the cluster reliabilities also are affected by the number of items in each cluster, with cluster scores based on fewer items having somewhat lower reliabilities than cluster scores based on more items.

Because the reliabilities of scores at the cluster level are lower, schools supplement the score results with other information when interpreting these results.

Subgroup reliabilities—The reliabilities of the operational STS for RLA are also examined for various subgroups of the examinee population that differed in their demographic characteristics. The characteristics considered are gender, economic status, provision of special services, length of attendance in U.S. schools, and EL program participation. The results of these analyses can be found in Table 8.B.2 through Table 8.B.11.

Reliability of performance classifications—The methodology used for estimating the reliability of classification decisions is described in the section “Decision Classification Analyses” on page 105. The results of these analyses are presented in Table 8.B.12 through Table 8.B.21 in Appendix 8.B; these tables start on page 133. When the classifications are collapsed to below-proficient versus proficient and above, the proportion of students that were classified accurately ranged from 0.91 to 0.94 across all the STS for RLA. Similarly, the proportion of students that were classified consistently ranged from 0.88 to 0.91 for students classified into below-proficient versus proficient and advanced.

These results represent high levels of decision accuracy and consistency.

Evidence Based on Consequences of Testing

As observed in the *Standards*, tests are usually administered “with the expectation that some benefit will be realized from the intended use of the scores” (AERA, APA, & NCME, 1999, p. 18). When this is the case, evidence that the expected benefits accrue will provide support for the intended use of the scores. The CDE and ETS are in the process of determining what kinds of information can be gathered to assess the consequences of administration of the STS for RLA.

IRT Analyses

Post-equating

Prior to the 2014–15 administration, the STS for RLA were equated to a reference form using a common-item nonequivalent groups design and post-equating methods based on IRT. The “base” or “reference” calibrations for the STS for RLA were established by calibrating samples of data from a specific administration. Doing so established a scale to which subsequent item calibrations could be linked.

The procedures used for post-equating the STS for RLA prior to 2012–13 involved three steps: item calibration, item parameter scaling, and production of raw-score-to-scale-score conversions using the scaled item parameters. ETS used GENASYM for the IRT item calibration and equating work. As part of this system, a proprietary version of the PARSCALE computer program (Muraki & Bock, 1995) was used and parameterized to result in one-parameter calibrations. Research at ETS has suggested that PARSCALE calibrations done in this manner produce results that are virtually identical to results based on WINSTEPS (Way, Kubiak, Henderson, & Julian, 2002). The post-equating procedures were applied to all the STS for RLA.

Pre-Equating

During the 2014–15 administration, because intact test forms from previous operational administrations were used without any edits or replacement of items, pre-equating was conducted prior to administration of the tests. Based on the sample invariant property of IRT, all the item parameter estimates were placed on the reference scale in their previous administrations through the post-equating procedure described above. For all STS for RLA forms, the conversion tables from previous administrations when the forms were originally used are directly applied to the current administration.

Descriptions of IRT analyses such as the model-data fit analyses can be found in Chapter 8 of the original-year technical report; the results of the IRT analyses are presented in Appendix 8.D of the original-year-technical report. *STS Technical Reports* are linked on the CDE’s Technical Reports and Studies Web page at <http://www.cde.ca.gov/ta/tg/sr/technicalrpts.asp>. The year of original administration for each multiple-choice STS for RLA is shown in Table 8.4.

The details on all equating procedures are given in Chapter 2, starting on page 8.

Summaries of Scaled IRT *b*-values

For the post-equating procedure prior to the 2013–14 administration, once the IRT *b*-values were placed on the item bank scale, analyses were performed to assess the overall test difficulty, the difficulty level of reporting clusters, and the distribution of items in a particular range of item difficulty.

During the 2014–15 administration, neither IRT calibrations nor scaling is implemented, but scaled *b*-value parameters derived through the post-equating procedure from their previous

administrations are used for pre-equating the STS for RLA. The summaries of b -values can be found in Appendix D of the *STS Technical Report* in the year the form was administered originally; see Table 8.4 on page 111 for administration years.

Evaluation of Pre-equating

Pre-equating is performed on the basis of the assumption of IRT models that item parameters remain invariant across samples given a similar ability distribution. To produce results that are sufficiently accurate for high-stakes decisions, previously administered operational forms were reused so that item parameters were obtained from large, representative samples, and factors that may affect item parameter estimations, such as context effects (e.g., item positions) and speededness, were well controlled.

To ensure that items performed similarly in the current administration as in the year they were originally administered in the intact forms, comparisons of classical statistics such as p -values and point-biserial correlations are made between the current administration and the item bank values in the year of the original administration.

Equating Results

During the 2014–15 administration, for all STS for RLA forms, the conversion tables from their original administrations (listed in Table 8.4 on page 111) are directly applied to the current administration.

Complete raw-score-to-scale-score conversion tables for the STS for RLA administered in 2014–15 are presented in *In 2014–15, administration of the STS for RLA was optional. For the majority of the grades, the number of students was lower than in previous administrations. To protect student privacy, frequency distribution is not shown if based on 10 or fewer student records. Because frequency distributions were not available for many raw score levels, no frequency distributions for any raw scores are included in Appendix 8.C.*

Table 8.C.1 through Table 8.C.10 starting on page 137. The raw scores and corresponding transformed scale scores are listed in those tables. The scale scores were truncated at both ends of the scale so that the minimum reported scale score was 150 and the maximum reported scale score was 600. The scale scores defining the various performance-level cut points are presented in Table 2.1, which is in Chapter 2 on page 16.

Differential Item Functioning Analyses

Analyses of DIF assess differences in the item performance of groups of students that differ in their demographic characteristics.

Prior to the 2013–14 administration, DIF analyses were performed based on the FIA sample, and were performed on all operational items and on all field-test items for which sufficient student samples were available. DIF analyses are not implemented during the 2014–15 administration because intact forms were used and all items were evaluated for DIF during the previous administration when the intact forms were originally used. These DIF results can be found in Appendix E of the *STS Technical Report* in the year the form was administered originally; see Table 8.4 on page 111 for administration years.

The statistical procedure of DIF analysis that was conducted prior to the 2014–15 administration is described in this section.

The sample size requirements for the DIF analyses were 100 in the focal group and 400 in the combined focal and reference groups. These sample sizes were based on standard

operating procedures with respect to DIF analyses at ETS. The DIF analyses utilized the Mantel-Haenszel (MH) DIF statistic (Mantel & Haenszel, 1959; Holland & Thayer, 1985). This statistic is based on the estimate of constant odds ratio and is described as the following:

The α_{MH} is the constant odds ratio taken from Dorans and Holland (1993, equation 7) and computed as the following:

$$\alpha_{MH} = \frac{\left(\sum_m R_{rm} \frac{W_{fm}}{N_{tm}} \right)}{\left(\sum_m R_{fm} \frac{W_{rm}}{N_{tm}} \right)} \quad (8.6)$$

$$MH\ D - DIF = -2.35 \ln[\alpha_{MH}] \quad (8.7)$$

where,

R = number right,

W = number wrong,

N = total in:

fm = focal group at ability m ,

rm = reference group at ability m , and

tm = total group at ability m .

Items analyzed for DIF at ETS are classified into one of three categories: A, B, or C. Category A contains items with negligible DIF. Category B contains items with slight to moderate DIF. Category C contains items with moderate to large values of DIF.

These categories have been used by ETS testing programs for more than 15 years. The definitions of the categories based on evaluations of the item-level MH D-DIF statistics are as follows:

DIF Category	Definition
A (negligible)	<ul style="list-style-type: none"> • Absolute value of MH D-DIF is not significantly different from zero, or is less than one. • Positive values are classified as “A+” and negative values as “A-.”
B (moderate)	<ul style="list-style-type: none"> • Absolute value of MH D-DIF is significantly different from zero but not from one, and is at least one; OR • Absolute value of MH D-DIF is significantly different from one, but is less than 1.5. • Positive values are classified as “B+” and negative values as “B-.”
C (large)	<ul style="list-style-type: none"> • Absolute value of MH D-DIF is significantly different from one, and is at least 1.5. • Positive values are classified as “C+” and negative values as “C-.”

The factors considered in the DIF analyses included gender and primary disability.

Tables also listed the operational and field-test items exhibiting significant DIF (C-DIF). Test developers were instructed to avoid selecting field-test items flagged as having shown C-DIF for future operational test forms unless their inclusion was deemed essential to meeting test-content specifications.

Tables showed the distributions of operational items across the DIF category classifications for the STS for RLA. In these tables, classifications of B- or C- indicated DIF against a focal group; classifications of B+ and C+ indicated DIF in favor of a focal group. The last two columns of each table showed the total number of items flagged for DIF in one or more comparisons.

References

- AERA, APA, & NCME. 2014. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- California Department of Education. (2011). *Standards-based Tests in Spanish technical report, spring 2011 administration*. Sacramento, CA. Retrieved from <http://www.cde.ca.gov/ta/tg/sr/documents/sts12techrpt.pdf>
- California Department of Education. (2012). *Standards-based Tests in Spanish technical report, spring 2012 administration*. Sacramento, CA. Retrieved from <http://www.cde.ca.gov/ta/tg/sr/documents/ststechrpta2011.pdf>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: Holt.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 292–334.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp.443–507). Washington, DC: American Council on Education.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York, NY: Wiley.
- Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Educational Testing Service. (2002). *ETS standards for quality and fairness*. Princeton, NJ: Author.
- Holland, P. W., & Thayer, D. T. (1985). *An alternative definition of the ETS delta scale of item difficulty* (Research Report 85–43). Princeton, NJ: Educational Testing Service.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp.17–64). Washington, DC: American Council on Education and National Council on Measurement in Education.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classification based on test scores. *Journal of Educational Measurement*, 32, 179–97.
- Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analyses of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719–48.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp.13–103). New York, NY: Macmillan.
- Muraki, E., & Bock, R. D. (1995). *PARSCALE: Parameter scaling of rating data* (Computer software, Version 2.2). Chicago, IL: Scientific Software.

United States Department of Education. (2001). Elementary and Secondary Education Act (Public Law 107–11), Title VI, Chapter B, § 4, Section 6162. Retrieved from <http://www2.ed.gov/policy/elsec/leg/esea02/index.html>

Way, W. D., Kubiak, A. T., Henderson, D., & Julian, M. W. (2002, April). *Accuracy and stability of calibrations for mixed-item-format tests using the 1-parameter and generalized partial credit models*. Paper presented at the meeting of the National Council on Measurement in Education, New Orleans, LA.

Appendix 8.A—Classical Analyses

Table 8.A.1 Item-by-item p -value and Point Biserial for RLA, Grades Two and Three—Current Year and Original Year of Administration

STS Year	2				3			
	2013		2015		2013		2015	
Items	p -value	Pt-Bis						
1	0.87	0.30	0.87	0.30	0.88	0.33	0.85	0.33
2	0.75	0.51	0.71	0.47	0.78	0.49	0.76	0.42
3	0.30	0.22	0.28	0.21	0.92	0.28	0.88	0.30
4	0.71	0.45	0.65	0.51	0.49	0.23	0.46	0.14
5	0.69	0.31	0.55	0.33	0.73	0.43	0.66	0.39
6	0.60	0.35	0.51	0.26	0.58	0.43	0.55	0.40
7	0.79	0.46	0.71	0.40	0.62	0.49	0.57	0.44
8	0.74	0.55	0.66	0.53	0.37	0.29	0.29	0.19
9	0.92	0.37	0.87	0.40	0.61	0.42	0.58	0.38
10	0.65	0.45	0.65	0.45	0.71	0.40	0.64	0.37
11	0.79	0.52	0.74	0.49	0.85	0.46	0.79	0.47
12	0.46	0.36	0.36	0.29	0.70	0.49	0.62	0.49
13	0.90	0.44	0.88	0.44	0.49	0.37	0.45	0.37
14	0.56	0.42	0.49	0.43	0.42	0.45	0.36	0.43
15	0.61	0.47	0.60	0.43	0.58	0.47	0.50	0.43
16	0.61	0.56	0.56	0.57	0.51	0.37	0.50	0.35
17	0.62	0.53	0.55	0.50	0.67	0.53	0.59	0.47
18	0.50	0.42	0.44	0.40	0.36	0.25	0.33	0.26
19	0.60	0.48	0.57	0.47	0.57	0.54	0.48	0.49
20	0.66	0.39	0.60	0.40	0.43	0.33	0.36	0.32
21	0.78	0.36	0.73	0.31	0.44	0.30	0.44	0.29
22	0.73	0.58	0.68	0.54	0.42	0.40	0.37	0.40
23	0.49	0.38	0.48	0.39	0.54	0.39	0.53	0.32
24	0.44	0.39	0.42	0.39	0.70	0.42	0.61	0.37
25	0.79	0.45	0.78	0.39	0.57	0.50	0.51	0.48
26	0.63	0.48	0.66	0.46	0.57	0.45	0.52	0.42
27	0.53	0.48	0.46	0.46	0.62	0.44	0.57	0.41
28	0.59	0.47	0.55	0.48	0.60	0.44	0.55	0.44
29	0.55	0.43	0.48	0.36	0.60	0.48	0.58	0.43
30	0.69	0.47	0.64	0.42	0.84	0.38	0.79	0.39
31	0.68	0.44	0.62	0.44	0.75	0.48	0.69	0.51
32	0.64	0.37	0.65	0.38	0.74	0.39	0.70	0.38
33	0.57	0.42	0.43	0.36	0.75	0.32	0.66	0.32
34	0.78	0.53	0.71	0.48	0.35	0.27	0.32	0.28
35	0.49	0.44	0.43	0.40	0.41	0.37	0.39	0.35
36	0.59	0.43	0.56	0.40	0.35	0.31	0.27	0.29
37	0.54	0.48	0.48	0.43	0.38	0.27	0.36	0.27
38	0.81	0.38	0.78	0.36	0.63	0.44	0.46	0.39
39	0.79	0.51	0.74	0.49	0.71	0.46	0.68	0.46
40	0.71	0.55	0.63	0.52	0.64	0.37	0.50	0.33
41	0.66	0.54	0.61	0.51	0.58	0.50	0.46	0.45
42	0.56	0.45	0.45	0.38	0.57	0.31	0.60	0.30
43	0.61	0.41	0.53	0.32	0.64	0.48	0.55	0.40
44	0.38	0.34	0.26	0.20	0.52	0.37	0.50	0.28
45	0.41	0.28	0.39	0.28	0.51	0.54	0.40	0.51
46	0.55	0.48	0.42	0.40	0.68	0.45	0.55	0.43
47	0.76	0.55	0.72	0.53	0.62	0.37	0.55	0.36
48	0.62	0.42	0.51	0.33	0.52	0.39	0.36	0.35
49	0.47	0.34	0.45	0.31	0.78	0.54	0.71	0.52

Year	2				3			
	2013		2015		2013		2015	
	p-value	Pt-Bis	p-value	Pt-Bis	p-value	Pt-Bis	p-value	Pt-Bis
50	0.44	0.49	0.34	0.39	0.59	0.33	0.53	0.18
51	0.69	0.54	0.63	0.54	0.75	0.52	0.57	0.42
52	0.46	0.34	0.43	0.27	0.68	0.42	0.40	0.26
53	0.60	0.50	0.47	0.46	0.45	0.35	0.37	0.23
54	0.52	0.46	0.46	0.35	0.45	0.34	0.37	0.26
55	0.60	0.46	0.55	0.43	0.57	0.39	0.54	0.32
56	0.60	0.59	0.41	0.49	0.59	0.54	0.51	0.51
57	0.57	0.47	0.53	0.46	0.75	0.55	0.57	0.51
58	0.65	0.52	0.61	0.55	0.61	0.38	0.57	0.34
59	0.58	0.41	0.56	0.39	0.53	0.55	0.43	0.50
60	0.66	0.48	0.64	0.43	0.52	0.30	0.46	0.23
61	0.54	0.55	0.53	0.54	0.43	0.28	0.35	0.25
62	0.55	0.48	0.48	0.44	0.77	0.50	0.68	0.44
63	0.47	0.21	0.48	0.25	0.34	0.23	0.33	0.18
64	0.69	0.36	0.61	0.34	0.46	0.40	0.36	0.33
65	0.51	0.52	0.44	0.46	0.38	0.23	0.36	0.22

Table 8.A.2 Item-by-item p -value and Point Biserial for RLA, Grades Four through Six—Current Year and Original Year of Administration

Year	4				5				6			
	2013		2015		2013		2015		2012		2015	
	p-value	Pt-Bis										
1	0.51	0.46	0.52	0.47	0.70	0.31	0.68	0.37	0.62	0.33	0.58	0.28
2	0.68	0.41	0.64	0.46	0.66	0.35	0.71	0.27	0.42	0.23	0.38	0.21
3	0.68	0.46	0.68	0.46	0.56	0.24	0.51	0.30	0.65	0.40	0.56	0.45
4	0.48	0.38	0.43	0.42	0.47	0.32	0.52	0.35	0.68	0.35	0.58	0.29
5	0.66	0.55	0.62	0.53	0.53	0.43	0.51	0.49	0.65	0.33	0.58	0.30
6	0.77	0.48	0.70	0.50	0.50	0.39	0.39	0.36	0.63	0.40	0.54	0.42
7	0.78	0.38	0.72	0.42	0.43	0.29	0.48	0.31	0.49	0.33	0.47	0.37
8	0.71	0.43	0.60	0.26	0.35	0.24	0.35	0.23	0.47	0.34	0.50	0.35
9	0.69	0.50	0.60	0.45	0.74	0.40	0.72	0.40	0.41	0.39	0.37	0.40
10	0.70	0.43	0.64	0.48	0.54	0.47	0.55	0.48	0.59	0.35	0.56	0.42
11	0.60	0.34	0.58	0.30	0.68	0.51	0.66	0.53	0.65	0.45	0.65	0.43
12	0.70	0.48	0.63	0.48	0.52	0.47	0.43	0.44	0.58	0.45	0.56	0.46
13	0.59	0.46	0.47	0.38	0.62	0.45	0.57	0.43	0.59	0.49	0.56	0.44
14	0.53	0.38	0.49	0.36	0.46	0.49	0.45	0.50	0.70	0.36	0.65	0.35
15	0.53	0.34	0.52	0.33	0.34	0.31	0.36	0.31	0.49	0.26	0.46	0.29
16	0.43	0.39	0.30	0.29	0.60	0.49	0.62	0.52	0.52	0.34	0.52	0.28
17	0.70	0.50	0.56	0.44	0.46	0.36	0.49	0.35	0.67	0.46	0.65	0.47
18	0.77	0.48	0.68	0.48	0.42	0.31	0.38	0.35	0.66	0.46	0.66	0.42
19	0.64	0.41	0.60	0.43	0.44	0.28	0.41	0.34	0.53	0.44	0.51	0.46
20	0.49	0.36	0.47	0.28	0.28	0.29	0.27	0.28	0.52	0.40	0.50	0.38
21	0.52	0.43	0.39	0.37	0.37	0.20	0.39	0.23	0.54	0.43	0.50	0.31
22	0.55	0.49	0.54	0.44	0.29	0.12	0.31	0.13	0.36	0.25	0.36	0.33
23	0.48	0.34	0.49	0.32	0.69	0.45	0.67	0.41	0.42	0.41	0.40	0.28
24	0.36	0.26	0.32	0.21	0.40	0.35	0.47	0.39	0.53	0.37	0.55	0.30
25	0.50	0.42	0.44	0.39	0.36	0.32	0.33	0.41	0.61	0.40	0.58	0.41
26	0.62	0.47	0.50	0.42	0.52	0.34	0.48	0.38	0.44	0.32	0.44	0.24
27	0.56	0.43	0.54	0.40	0.44	0.40	0.44	0.40	0.56	0.42	0.52	0.35
28	0.67	0.57	0.61	0.57	0.56	0.46	0.57	0.47	0.42	0.24	0.39	0.28
29	0.73	0.54	0.67	0.52	0.47	0.35	0.45	0.34	0.56	0.48	0.57	0.52
30	0.65	0.56	0.52	0.49	0.49	0.37	0.44	0.32	0.45	0.27	0.41	0.28
31	0.42	0.32	0.34	0.20	0.38	0.28	0.34	0.24	0.51	0.36	0.43	0.39

STS	4				5				6			
Year	2013		2015		2013		2015		2012		2015	
Items	p-value	Pt-Bis										
32	0.35	0.25	0.30	0.18	0.37	0.33	0.33	0.33	0.30	0.22	0.28	0.30
33	0.43	0.27	0.38	0.32	0.49	0.45	0.50	0.46	0.43	0.35	0.41	0.24
34	0.44	0.44	0.41	0.38	0.53	0.49	0.53	0.51	0.45	0.28	0.41	0.25
35	0.39	0.23	0.29	0.04	0.54	0.50	0.54	0.52	0.37	0.27	0.42	0.29
36	0.58	0.28	0.53	0.28	0.53	0.40	0.54	0.39	0.26	0.13	0.27	0.15
37	0.82	0.45	0.75	0.47	0.49	0.41	0.47	0.43	0.41	0.29	0.40	0.25
38	0.87	0.46	0.82	0.44	0.67	0.48	0.63	0.49	0.63	0.39	0.62	0.39
39	0.73	0.49	0.64	0.56	0.49	0.37	0.47	0.37	0.46	0.36	0.51	0.37
40	0.63	0.41	0.58	0.32	0.81	0.39	0.76	0.40	0.49	0.51	0.44	0.60
41	0.72	0.41	0.66	0.51	0.71	0.31	0.72	0.36	0.67	0.31	0.60	0.34
42	0.76	0.45	0.69	0.47	0.72	0.44	0.69	0.54	0.56	0.42	0.55	0.47
43	0.49	0.35	0.40	0.27	0.61	0.44	0.58	0.50	0.48	0.40	0.48	0.31
44	0.68	0.52	0.62	0.51	0.52	0.36	0.51	0.40	0.51	0.38	0.46	0.43
45	0.61	0.47	0.57	0.46	0.53	0.31	0.54	0.35	0.47	0.31	0.48	0.33
46	0.48	0.37	0.43	0.36	0.71	0.46	0.67	0.52	0.41	0.30	0.36	0.32
47	0.30	0.25	0.27	0.28	0.45	0.32	0.45	0.32	0.86	0.43	0.73	0.46
48	0.38	0.33	0.33	0.23	0.47	0.21	0.50	0.31	0.52	0.33	0.42	0.35
49	0.59	0.43	0.54	0.44	0.56	0.30	0.55	0.37	0.36	0.38	0.29	0.37
50	0.71	0.53	0.68	0.54	0.52	0.40	0.52	0.42	0.62	0.42	0.54	0.38
51	0.53	0.47	0.47	0.48	0.45	0.33	0.44	0.32	0.51	0.32	0.43	0.35
52	0.70	0.40	0.68	0.41	0.52	0.29	0.46	0.29	0.66	0.51	0.59	0.51
53	0.54	0.54	0.47	0.53	0.68	0.42	0.69	0.45	0.47	0.32	0.35	0.30
54	0.47	0.39	0.43	0.41	0.31	0.26	0.32	0.25	0.57	0.38	0.51	0.30
55	0.55	0.46	0.48	0.37	0.54	0.34	0.54	0.34	0.55	0.42	0.44	0.41
56	0.55	0.49	0.49	0.52	0.60	0.42	0.59	0.35	0.52	0.36	0.48	0.45
57	0.63	0.51	0.53	0.47	0.47	0.30	0.45	0.31	0.45	0.29	0.31	0.24
58	0.35	0.32	0.30	0.25	0.63	0.46	0.60	0.47	0.52	0.52	0.44	0.47
59	0.59	0.26	0.51	0.25	0.59	0.50	0.56	0.55	0.40	0.36	0.35	0.37
60	0.36	0.32	0.30	0.29	0.54	0.40	0.52	0.40	0.40	0.29	0.37	0.34
61	0.70	0.49	0.68	0.47	0.48	0.40	0.50	0.41	0.45	0.35	0.39	0.27
62	0.43	0.34	0.39	0.33	0.38	0.40	0.33	0.47	0.65	0.50	0.59	0.47
63	0.70	0.54	0.60	0.52	0.47	0.35	0.45	0.41	0.55	0.38	0.40	0.41
64	0.37	0.36	0.34	0.29	0.51	0.35	0.51	0.33	0.63	0.45	0.52	0.44
65	0.55	0.42	0.50	0.39	0.28	0.28	0.27	0.28	0.42	0.32	0.42	0.34
66	0.51	0.40	0.49	0.37	0.51	0.47	0.51	0.46	0.64	0.46	0.55	0.47
67	0.35	0.25	0.30	0.20	0.34	0.28	0.30	0.29	0.54	0.32	0.54	0.34
68	0.55	0.33	0.50	0.35	0.61	0.50	0.60	0.50	0.49	0.34	0.43	0.35
69	0.59	0.42	0.56	0.41	0.54	0.43	0.53	0.44	0.58	0.51	0.54	0.53
70	0.63	0.43	0.56	0.37	0.42	0.40	0.36	0.37	0.66	0.48	0.62	0.53
71	0.61	0.47	0.57	0.47	0.54	0.35	0.55	0.39	0.32	0.23	0.25	0.16
72	0.69	0.54	0.66	0.48	0.48	0.38	0.42	0.34	0.57	0.51	0.55	0.53
73	0.43	0.35	0.40	0.38	0.46	0.37	0.47	0.39	0.39	0.23	0.32	0.21
74	0.55	0.32	0.50	0.36	0.31	0.30	0.30	0.30	0.50	0.35	0.45	0.22
75	0.63	0.48	0.61	0.50	0.41	0.33	0.37	0.36	0.32	0.29	0.28	0.27

Table 8.A.3 Item-by-item p -value and Point Biserial for RLA, Grades Seven through Nine—Current Year and Original Year of Administration

STS	7				8				9			
	Year	2013	2015	2012	2015	2011	2015	2011	2015			
Items	p -value	Pt-Bis										
1	0.60	0.22	0.57	0.33	0.38	0.31	0.48	0.32	0.29	0.19	0.27	0.22
2	0.74	0.44	0.72	0.47	0.32	0.24	0.39	0.25	0.37	0.22	0.38	0.21
3	0.73	0.27	0.74	0.32	0.33	0.28	0.39	0.23	0.72	0.25	0.70	0.26
4	0.53	0.32	0.51	0.29	0.50	0.41	0.64	0.32	0.37	0.30	0.40	0.17
5	0.60	0.43	0.56	0.37	0.40	0.24	0.45	0.36	0.43	0.30	0.37	0.22
6	0.74	0.41	0.65	0.44	0.48	0.32	0.47	0.34	0.27	0.36	0.18	0.35
7	0.32	0.34	0.26	0.35	0.53	0.46	0.53	0.53	0.84	0.35	0.84	0.35
8	0.79	0.41	0.69	0.41	0.73	0.39	0.70	0.42	0.75	0.39	0.76	0.43
9	0.77	0.42	0.71	0.39	0.74	0.44	0.66	0.50	0.61	0.30	0.52	0.25
10	0.48	0.42	0.40	0.38	0.48	0.37	0.46	0.42	0.43	0.15	0.41	0.19
11	0.47	0.31	0.41	0.30	0.49	0.23	0.48	0.18	0.83	0.45	0.81	0.45
12	0.69	0.39	0.64	0.43	0.72	0.30	0.68	0.35	0.76	0.44	0.70	0.43
13	0.38	0.44	0.34	0.40	0.70	0.31	0.63	0.33	0.82	0.51	0.78	0.51
14	0.50	0.46	0.42	0.46	0.58	0.25	0.52	0.29	0.49	0.26	0.49	0.32
15	0.59	0.37	0.55	0.33	0.45	0.26	0.45	0.24	0.47	0.33	0.45	0.28
16	0.56	0.35	0.57	0.31	0.42	0.38	0.44	0.40	0.41	0.28	0.34	0.18
17	0.76	0.44	0.74	0.45	0.53	0.37	0.51	0.39	0.80	0.44	0.79	0.37
18	0.56	0.46	0.51	0.43	0.56	0.49	0.54	0.46	0.70	0.47	0.76	0.40
19	0.53	0.33	0.52	0.35	0.34	0.21	0.34	0.16	0.66	0.33	0.71	0.30
20	0.40	0.30	0.35	0.32	0.74	0.43	0.72	0.43	0.72	0.55	0.74	0.46
21	0.54	0.35	0.49	0.41	0.42	0.48	0.44	0.43	0.61	0.36	0.59	0.34
22	0.49	0.27	0.47	0.26	0.63	0.49	0.57	0.46	0.83	0.51	0.86	0.40
23	0.42	0.40	0.37	0.36	0.46	0.29	0.42	0.33	0.43	0.27	0.42	0.26
24	0.67	0.46	0.68	0.46	0.37	0.38	0.36	0.39	0.80	0.46	0.86	0.39
25	0.55	0.50	0.45	0.49	0.66	0.39	0.59	0.46	0.57	0.45	0.56	0.37
26	0.59	0.45	0.56	0.48	0.60	0.54	0.52	0.53	0.54	0.46	0.50	0.47
27	0.42	0.32	0.42	0.28	0.43	0.30	0.42	0.19	0.60	0.41	0.64	0.34
28	0.72	0.51	0.64	0.45	0.71	0.41	0.65	0.44	0.55	0.35	0.52	0.33
29	0.49	0.37	0.45	0.34	0.49	0.42	0.44	0.35	0.39	0.31	0.36	0.31
30	0.49	0.34	0.48	0.41	0.60	0.48	0.54	0.42	0.74	0.45	0.72	0.42
31	0.47	0.38	0.41	0.39	0.54	0.32	0.51	0.36	0.43	0.34	0.48	0.30
32	0.60	0.49	0.55	0.46	0.28	0.28	0.25	0.21	0.64	0.49	0.57	0.48
33	0.63	0.29	0.68	0.36	0.33	0.32	0.27	0.25	0.48	0.31	0.42	0.37
34	0.59	0.50	0.51	0.44	0.35	0.33	0.33	0.32	0.72	0.44	0.72	0.40
35	0.35	0.20	0.36	0.22	0.51	0.45	0.46	0.51	0.55	0.39	0.55	0.41
36	0.47	0.39	0.47	0.34	0.45	0.43	0.44	0.50	0.64	0.45	0.62	0.41
37	0.67	0.49	0.58	0.49	0.49	0.40	0.44	0.39	0.61	0.52	0.61	0.49
38	0.48	0.29	0.44	0.33	0.36	0.35	0.33	0.21	0.47	0.33	0.42	0.32
39	0.59	0.39	0.54	0.35	0.48	0.36	0.42	0.39	0.70	0.45	0.68	0.40
40	0.72	0.37	0.65	0.33	0.78	0.45	0.70	0.48	0.40	0.34	0.34	0.31
41	0.78	0.40	0.75	0.40	0.27	0.13	0.29	0.14	0.41	0.26	0.40	0.32
42	0.57	0.44	0.47	0.42	0.61	0.26	0.63	0.29	0.46	0.35	0.39	0.36
43	0.56	0.36	0.51	0.41	0.71	0.38	0.63	0.42	0.35	0.16	0.33	0.14
44	0.57	0.44	0.51	0.43	0.44	0.29	0.38	0.25	0.49	0.34	0.49	0.37
45	0.70	0.36	0.61	0.38	0.52	0.25	0.49	0.26	0.36	0.23	0.33	0.18
46	0.50	0.38	0.43	0.34	0.42	0.22	0.44	0.21	0.49	0.27	0.47	0.23
47	0.71	0.39	0.66	0.43	0.84	0.43	0.79	0.44	0.41	0.36	0.36	0.33
48	0.79	0.42	0.70	0.47	0.52	0.39	0.51	0.40	0.63	0.48	0.56	0.46
49	0.69	0.34	0.67	0.40	0.39	0.28	0.38	0.20	0.79	0.41	0.71	0.37
50	0.64	0.50	0.54	0.52	0.53	0.29	0.49	0.22	0.59	0.36	0.53	0.38
51	0.73	0.43	0.66	0.50	0.53	0.34	0.44	0.34	0.66	0.37	0.60	0.36
52	0.68	0.52	0.58	0.49	0.85	0.39	0.80	0.45	0.48	0.31	0.45	0.23

STS		7				8				9			
Year	2013	2015		2012		2015		2011		2015			
Items	<i>p</i> -value	Pt-Bis											
53	0.66	0.42	0.62	0.41	0.53	0.34	0.49	0.33	0.71	0.41	0.71	0.34	
54	0.75	0.45	0.67	0.50	0.39	0.37	0.30	0.34	0.35	0.28	0.33	0.28	
55	0.40	0.31	0.34	0.21	0.51	0.46	0.49	0.42	0.61	0.31	0.57	0.32	
56	0.52	0.48	0.41	0.57	0.78	0.44	0.78	0.38	0.48	0.41	0.40	0.35	
57	0.50	0.47	0.48	0.46	0.76	0.38	0.67	0.39	0.62	0.44	0.61	0.47	
58	0.35	0.25	0.29	0.28	0.63	0.38	0.58	0.38	0.51	0.33	0.48	0.43	
59	0.48	0.47	0.44	0.53	0.62	0.51	0.52	0.52	0.68	0.33	0.66	0.32	
60	0.55	0.46	0.49	0.43	0.68	0.43	0.60	0.44	0.26	0.26	0.25	0.25	
61	0.53	0.35	0.53	0.40	0.62	0.50	0.56	0.48	0.49	0.47	0.45	0.48	
62	0.56	0.41	0.47	0.44	0.46	0.35	0.39	0.35	0.49	0.37	0.45	0.34	
63	0.63	0.45	0.62	0.45	0.61	0.40	0.57	0.32	0.24	0.08	0.21	0.05	
64	0.66	0.50	0.57	0.49	0.46	0.33	0.43	0.39	0.77	0.41	0.75	0.35	
65	0.44	0.43	0.45	0.40	0.54	0.39	0.49	0.36	0.72	0.53	0.72	0.45	
66	0.72	0.40	0.69	0.41	0.69	0.46	0.61	0.44	0.67	0.39	0.61	0.35	
67	0.78	0.34	0.73	0.42	0.56	0.36	0.48	0.41	0.46	0.40	0.41	0.36	
68	0.61	0.49	0.58	0.49	0.38	0.37	0.33	0.29	0.35	0.23	0.33	0.19	
69	0.38	0.27	0.34	0.26	0.46	0.40	0.41	0.38	0.65	0.45	0.60	0.44	
70	0.43	0.29	0.32	0.29	0.43	0.30	0.32	0.31	0.64	0.51	0.56	0.47	
71	0.50	0.41	0.47	0.39	0.49	0.40	0.48	0.42	0.31	0.36	0.29	0.36	
72	0.58	0.45	0.53	0.39	0.43	0.29	0.38	0.30	0.73	0.46	0.69	0.49	
73	0.62	0.30	0.55	0.43	0.41	0.30	0.35	0.34	0.46	0.41	0.47	0.41	
74	0.54	0.42	0.51	0.53	0.52	0.38	0.47	0.42	0.34	0.29	0.31	0.33	
75	0.34	0.25	0.27	0.09	0.80	0.40	0.76	0.41	0.44	0.39	0.41	0.34	

Table 8.A.4 Item-by-item *p*-value and Point Biserial for RLA, Grades Ten and Eleven—Current Year and Original Year of Administration

STS		10				11			
Year	2013	2015		2013		2015			
Items	<i>p</i> -value	Pt-Bis							
1	0.59	0.32	0.56	0.32	0.42	0.31	0.37	0.39	
2	0.77	0.31	0.72	0.33	0.63	0.44	0.63	0.37	
3	0.55	0.39	0.46	0.32	0.68	0.36	0.67	0.39	
4	0.78	0.30	0.74	0.23	0.55	0.37	0.50	0.36	
5	0.63	0.38	0.58	0.31	0.44	0.21	0.43	0.34	
6	0.66	0.41	0.67	0.40	0.64	0.45	0.65	0.45	
7	0.44	0.32	0.45	0.29	0.72	0.30	0.66	0.31	
8	0.47	0.38	0.40	0.42	0.24	0.07	0.23	0.10	
9	0.67	0.40	0.64	0.38	0.43	0.32	0.48	0.24	
10	0.70	0.28	0.67	0.34	0.69	0.48	0.66	0.52	
11	0.71	0.51	0.74	0.42	0.76	0.06	0.76	0.17	
12	0.54	0.34	0.48	0.45	0.62	0.35	0.64	0.48	
13	0.56	0.29	0.50	0.45	0.51	0.01	0.54	0.07	
14	0.81	0.39	0.75	0.51	0.66	0.35	0.62	0.51	
15	0.73	0.34	0.69	0.43	0.58	0.32	0.61	0.34	
16	0.51	0.26	0.48	0.23	0.72	0.44	0.67	0.50	
17	0.84	0.45	0.78	0.59	0.68	0.36	0.73	0.25	
18	0.32	0.33	0.34	0.47	0.36	0.26	0.39	0.38	
19	0.45	0.25	0.42	0.32	0.41	0.26	0.40	0.26	
20	0.77	0.44	0.74	0.45	0.80	0.43	0.76	0.57	
21	0.68	0.40	0.61	0.48	0.37	0.44	0.38	0.47	
22	0.65	0.47	0.60	0.53	0.55	0.42	0.63	0.37	
23	0.69	0.43	0.63	0.44	0.57	0.39	0.60	0.51	
24	0.71	0.41	0.71	0.30	0.51	0.42	0.54	0.27	

STS	10				11			
	Year	2013		2015		2013		2015
Items	p-value	Pt-Bis	p-value	Pt-Bis	p-value	Pt-Bis	p-value	Pt-Bis
25	0.60	0.39	0.55	0.34	0.38	0.33	0.37	0.36
26	0.71	0.39	0.62	0.45	0.39	0.16	0.41	0.25
27	0.51	0.31	0.48	0.35	0.45	0.15	0.40	0.12
28	0.88	0.42	0.84	0.52	0.51	0.38	0.52	0.42
29	0.45	0.22	0.46	0.14	0.36	0.20	0.45	0.23
30	0.36	0.32	0.33	0.23	0.50	0.21	0.56	0.19
31	0.53	0.36	0.54	0.35	0.59	0.52	0.62	0.53
32	0.41	0.21	0.42	0.23	0.52	0.10	0.44	0.17
33	0.68	0.41	0.69	0.48	0.33	0.17	0.34	-0.02
34	0.58	0.32	0.53	0.32	0.61	0.37	0.54	0.52
35	0.61	0.44	0.63	0.35	0.46	0.47	0.44	0.52
36	0.57	0.37	0.55	0.42	0.72	0.46	0.64	0.47
37	0.42	0.36	0.44	0.29	0.51	0.33	0.47	0.35
38	0.80	0.50	0.75	0.62	0.41	0.34	0.44	0.49
39	0.64	0.44	0.55	0.39	0.72	0.24	0.76	0.30
40	0.38	0.30	0.41	0.26	0.70	0.37	0.68	0.42
41	0.70	0.38	0.65	0.45	0.61	0.34	0.60	0.38
42	0.55	0.23	0.55	0.34	0.64	0.40	0.60	0.39
43	0.60	0.46	0.58	0.45	0.88	0.45	0.81	0.49
44	0.51	0.37	0.49	0.38	0.37	0.30	0.41	0.13
45	0.50	0.43	0.48	0.48	0.51	0.38	0.52	0.51
46	0.66	0.47	0.57	0.39	0.59	0.41	0.53	0.38
47	0.37	0.29	0.40	0.30	0.62	0.41	0.60	0.39
48	0.52	0.31	0.45	0.36	0.51	0.46	0.51	0.56
49	0.39	0.20	0.37	0.28	0.53	0.43	0.50	0.52
50	0.51	0.22	0.47	0.21	0.57	0.30	0.56	0.32
51	0.42	0.31	0.37	0.32	0.44	0.25	0.43	0.25
52	0.55	0.42	0.48	0.45	0.55	0.38	0.47	0.42
53	0.33	0.24	0.29	0.33	0.43	0.28	0.40	0.37
54	0.51	0.39	0.45	0.41	0.61	0.38	0.59	0.32
55	0.56	0.36	0.46	0.38	0.46	0.46	0.36	0.58
56	0.38	0.35	0.33	0.39	0.28	0.27	0.23	0.42
57	0.80	0.30	0.80	0.32	0.64	0.44	0.57	0.54
58	0.65	0.37	0.63	0.42	0.43	0.45	0.40	0.40
59	0.56	0.44	0.52	0.28	0.71	0.41	0.66	0.49
60	0.36	0.28	0.31	0.32	0.64	0.42	0.71	0.47
61	0.35	0.18	0.33	0.16	0.75	0.33	0.78	0.19
62	0.82	0.49	0.78	0.54	0.58	0.39	0.60	0.39
63	0.72	0.50	0.68	0.49	0.59	0.44	0.60	0.35
64	0.44	0.29	0.48	0.24	0.72	0.47	0.66	0.47
65	0.63	0.27	0.64	0.40	0.51	0.35	0.54	0.28
66	0.30	0.26	0.28	0.07	0.37	0.34	0.37	0.16
67	0.61	0.45	0.58	0.41	0.39	0.16	0.42	0.27
68	0.64	0.47	0.56	0.51	0.73	0.45	0.67	0.47
69	0.41	0.30	0.37	0.33	0.49	0.38	0.50	0.40
70	0.59	0.38	0.52	0.24	0.51	0.46	0.48	0.50
71	0.51	0.35	0.49	0.33	0.52	0.38	0.50	0.38
72	0.58	0.26	0.54	0.38	0.71	0.42	0.68	0.37
73	0.30	0.39	0.22	0.45	0.60	0.34	0.63	0.35
74	0.70	0.37	0.64	0.41	0.61	0.37	0.55	0.49
75	0.47	0.25	0.41	0.17	0.84	0.45	0.85	0.48

Appendix 8.B—Reliability Analyses

The reliabilities are reported only for samples that comprise 11 or more examinees. Also, in some cases in Appendix 8.B, score reliabilities were not estimable and are presented in the tables as hyphens. Finally, results based on samples that contain 50 or fewer examinees should be interpreted with caution due to small sample sizes.

Table 8.B.1 Subscore Reliabilities and Intercorrelations for RLA

Subscore Area	No. of Items	Intercorrelation					Reliab.	SEM
		1	2	3	4	5		
Grade 2								
1. Word Analysis and Vocabulary Development	22	1.00	0.81	1.97
2. Reading Comprehension	15	0.73	1.00	.	.	.	0.77	1.70
3. Literary Response and Analysis	6	0.58	0.58	1.00	.	.	0.56	1.06
4. Written Conventions	14	0.75	0.71	0.58	1.00	.	0.78	1.64
5. Writing Strategies	8	0.53	0.54	0.47	0.57	1.00	0.52	1.29
Grade 3								
1. Word Analysis and Vocabulary Development	20	1.00	0.76	1.94
2. Reading Comprehension	15	0.67	1.00	.	.	.	0.72	1.76
3. Literary Response and Analysis	8	0.56	0.55	1.00	.	.	0.44	1.25
4. Written Conventions	13	0.62	0.59	0.49	1.00	.	0.68	1.64
5. Writing Strategies	9	0.55	0.57	0.45	0.57	1.00	0.56	1.38
Grade 4								
1. Word Analysis and Vocabulary Development	18	1.00	0.78	1.87
2. Reading Comprehension	15	0.70	1.00	.	.	.	0.74	1.75
3. Literary Response and Analysis	9	0.58	0.62	1.00	.	.	0.56	1.36
4. Written Conventions	18	0.73	0.66	0.59	1.00	.	0.80	1.83
5. Writing Strategies	15	0.68	0.65	0.58	0.69	1.00	0.68	1.75
Grade 5								
1. Word Analysis and Vocabulary Development	14	1.00	0.71	1.67
2. Reading Comprehension	16	0.69	1.00	.	.	.	0.67	1.83
3. Literary Response and Analysis	12	0.70	0.66	1.00	.	.	0.70	1.55
4. Written Conventions	17	0.67	0.62	0.66	1.00	.	0.77	1.83
5. Writing Strategies	16	0.68	0.63	0.67	0.70	1.00	0.71	1.81
Grade 6								
1. Word Analysis and Vocabulary Development	13	1.00	0.70	1.62
2. Reading Comprehension	17	0.61	1.00	.	.	.	0.69	1.91
3. Literary Response and Analysis	12	0.59	0.61	1.00	.	.	0.63	1.60
4. Written Conventions	16	0.59	0.59	0.58	1.00	.	0.75	1.80
5. Writing Strategies	17	0.64	0.60	0.53	0.66	1.00	0.71	1.87
Grade 7								
1. Word Analysis and Vocabulary Development	11	1.00	0.70	1.46
2. Reading Comprehension	18	0.68	1.00	.	.	.	0.75	1.90
3. Literary Response and Analysis	13	0.67	0.67	1.00	.	.	0.71	1.61
4. Written Conventions	16	0.64	0.65	0.57	1.00	.	0.77	1.77
5. Writing Strategies	17	0.68	0.67	0.63	0.75	1.00	0.76	1.83
Grade 8								
1. Word Analysis and Vocabulary Development	9	1.00	0.64	1.34
2. Reading Comprehension	18	0.61	1.00	.	.	.	0.73	1.92
3. Literary Response and Analysis	15	0.64	0.60	1.00	.	.	0.70	1.77
4. Written Conventions	16	0.61	0.59	0.56	1.00	.	0.73	1.76

5. Writing Strategies	17	0.60	0.55	0.57	0.64	1.00	0.68	1.88
Grade 9		1	2	3	4	5		
1. Word Analysis and Vocabulary Development	8	1.00	0.59	1.14
2. Reading Comprehension	18	0.63	1.00	.	.	.	0.73	1.86
3. Literary Response and Analysis	16	0.57	0.64	1.00	.	.	0.62	1.77
4. Written Conventions	13	0.55	0.59	0.50	1.00	.	0.64	1.62
5. Writing Strategies	20	0.51	0.63	0.51	0.63	1.00	0.72	2.03
Grade 10		1	2	3	4	5		
1. Word Analysis and Vocabulary Development	8	1.00	0.68	1.15
2. Reading Comprehension	18	0.67	1.00	.	.	.	0.75	1.91
3. Literary Response and Analysis	16	0.64	0.66	1.00	.	.	0.69	1.78
4. Written Conventions	13	0.67	0.67	0.58	1.00	.	0.72	1.59
5. Writing Strategies	20	0.62	0.62	0.57	0.64	1.00	0.67	2.07
Grade 11		1	2	3	4	5		
1. Word Analysis and Vocabulary Development	8	1.00	0.51	1.28
2. Reading Comprehension	19	0.56	1.00	.	.	.	0.54	2.05
3. Literary Response and Analysis	17	0.66	0.66	1.00	.	.	0.78	1.81
4. Written Conventions	9	0.53	0.55	0.57	1.00	.	0.59	1.28
5. Writing Strategies	22	0.66	0.66	0.72	0.65	1.00	0.82	2.08

Table 8.B.2 Reliabilities and SEMs by Gender (Male)

STS	N	Reliab.	SEM
2 RLA	946	0.92	3.56
3 RLA	815	0.89	3.66
4 RLA	473	0.93	3.89
5 RLA	390	0.92	3.91
6 RLA	240	0.90	3.98
7 RLA	221	0.92	3.91
8 RLA	243	0.91	3.94
9 RLA	468	0.90	3.87
10 RLA	168	0.91	3.92
11 RLA	65	0.91	3.96

Table 8.B.3 Reliabilities and SEMs by Gender (Female)

STS	N	Reliab.	SEM
2 RLA	933	0.92	3.50
3 RLA	828	0.90	3.60
4 RLA	461	0.92	3.85
5 RLA	340	0.92	3.92
6 RLA	200	0.91	3.96
7 RLA	234	0.93	3.84
8 RLA	214	0.90	3.93
9 RLA	342	0.89	3.83
10 RLA	117	0.91	3.86
11 RLA	66	0.90	3.82

Table 8.B.4 Reliabilities and SEMs by Economic Status (Not Disadvantaged)

STS	N	Reliab.	SEM
2 RLA	158	0.92	3.57
3 RLA	164	0.92	3.57
4 RLA	142	0.93	3.84
5 RLA	123	0.93	3.87
6 RLA	105	0.92	3.93
7 RLA	98	0.93	3.88
8 RLA	113	0.92	3.91
9 RLA	273	0.90	3.84
10 RLA	61	0.91	3.86
11 RLA	22	0.94	3.84

Table 8.B.5 Reliabilities and SEMs by Economic Status (Disadvantaged)

STS	N	Reliab.	SEM
2 RLA	1,721	0.92	3.53
3 RLA	1,479	0.90	3.64
4 RLA	792	0.93	3.88
5 RLA	607	0.92	3.92
6 RLA	335	0.91	3.99
7 RLA	357	0.93	3.88
8 RLA	344	0.90	3.95
9 RLA	537	0.90	3.86
10 RLA	224	0.91	3.90
11 RLA	109	0.91	3.91

Table 8.B.6 Reliabilities and SEMs by Special Services (No Special Services)

STS	N	Reliab.	SEM
2 RLA	1,794	0.92	3.53
3 RLA	1,563	0.90	3.63
4 RLA	890	0.93	3.87
5 RLA	698	0.92	3.92
6 RLA	435	0.91	3.97
7 RLA	450	0.93	3.88
8 RLA	449	0.91	3.94
9 RLA	809	0.90	3.85
10 RLA	284	0.91	3.89
11 RLA	131	0.91	3.90

Table 8.B.7 Reliabilities and SEMs by Special Services (Special Services)

STS	N	Reliab.	SEM
2 RLA	85	0.92	3.59
3 RLA	80	0.86	3.68
4 RLA	44	0.89	3.88
5 RLA	32	0.87	3.81
6 RLA	5	–	–
7 RLA	5	–	–
8 RLA	8	–	–
9 RLA	1	–	–
10 RLA	1	–	–
11 RLA	0	–	–

Table 8.B.8 Reliabilities and SEMs by Attendance in U.S. Schools (< 12 Months)

STS	N	Reliab.	SEM
2 RLA	250	0.92	3.59
3 RLA	306	0.91	3.59
4 RLA	319	0.92	3.89
5 RLA	270	0.93	3.88
6 RLA	273	0.91	3.96
7 RLA	323	0.92	3.87
8 RLA	333	0.91	3.91
9 RLA	703	0.90	3.85
10 RLA	194	0.91	3.86
11 RLA	97	0.91	3.88

Table 8.B.9 Reliabilities and SEMs by Attendance in U.S. Schools (≥ 12 Months)

STS	N	Reliab.	SEM
2 RLA	1,629	0.92	3.52
3 RLA	1,337	0.89	3.63
4 RLA	615	0.93	3.86
5 RLA	460	0.92	3.92
6 RLA	167	0.91	3.99
7 RLA	132	0.93	3.88
8 RLA	124	0.89	4.00
9 RLA	107	0.89	3.91
10 RLA	91	0.91	3.94
11 RLA	34	0.92	3.96

Table 8.B.10 Subscore Reliabilities and SEM for RLA by Gender/Economic Status

Subscore Area	No. of Items	Male		Female		Not Econ. Dis.		Econ. Dis.	
		Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
Grade 2									
1. Word Analysis and Vocabulary Development	22	0.81	2.00	0.80	1.94	0.80	2.03	0.81	1.97
2. Reading Comprehension	15	0.77	1.71	0.77	1.68	0.77	1.71	0.77	1.69
3. Literary Response and Analysis	6	0.55	1.07	0.56	1.05	0.55	1.08	0.56	1.06
4. Written Conventions	14	0.77	1.65	0.78	1.63	0.79	1.64	0.77	1.64
5. Writing Strategies	8	0.54	1.28	0.51	1.30	0.53	1.26	0.52	1.29
Grade 3									
1. Word Analysis and Vocabulary Development	20	0.73	1.97	0.78	1.91	0.80	1.89	0.75	1.95
2. Reading Comprehension	15	0.68	1.77	0.74	1.74	0.73	1.74	0.71	1.76
3. Literary Response and Analysis	8	0.38	1.26	0.48	1.24	0.50	1.24	0.43	1.25
4. Written Conventions	13	0.66	1.65	0.68	1.63	0.77	1.59	0.66	1.65
5. Writing Strategies	9	0.53	1.39	0.58	1.37	0.65	1.35	0.54	1.38
Grade 4									
1. Word Analysis and Vocabulary Development	18	0.78	1.89	0.77	1.84	0.77	1.86	0.78	1.87
2. Reading Comprehension	15	0.74	1.75	0.72	1.74	0.71	1.74	0.74	1.75
3. Literary Response and Analysis	9	0.56	1.36	0.54	1.36	0.56	1.37	0.57	1.36
4. Written Conventions	18	0.81	1.85	0.78	1.80	0.82	1.81	0.80	1.84
5. Writing Strategies	15	0.69	1.74	0.66	1.76	0.71	1.74	0.68	1.75
Grade 5									
1. Word Analysis and Vocabulary Development	14	0.71	1.68	0.70	1.67	0.70	1.68	0.71	1.67
2. Reading Comprehension	16	0.67	1.82	0.66	1.83	0.69	1.83	0.66	1.83
3. Literary Response and Analysis	12	0.70	1.56	0.69	1.54	0.76	1.51	0.69	1.56
4. Written Conventions	17	0.76	1.84	0.76	1.81	0.83	1.76	0.76	1.84
5. Writing Strategies	16	0.73	1.79	0.68	1.83	0.74	1.79	0.71	1.81
Grade 6									
1. Word Analysis and Vocabulary Development	13	0.67	1.62	0.71	1.62	0.68	1.62	0.70	1.62
2. Reading Comprehension	17	0.70	1.90	0.65	1.92	0.71	1.89	0.68	1.92
3. Literary Response and Analysis	12	0.59	1.62	0.65	1.56	0.69	1.55	0.60	1.61
4. Written Conventions	16	0.75	1.81	0.75	1.78	0.78	1.77	0.74	1.81
5. Writing Strategies	17	0.65	1.86	0.73	1.87	0.73	1.86	0.70	1.87
Grade 7									
1. Word Analysis and Vocabulary Development	11	0.70	1.48	0.69	1.44	0.69	1.47	0.70	1.46
2. Reading Comprehension	18	0.73	1.92	0.75	1.88	0.73	1.91	0.75	1.90
3. Literary Response and Analysis	13	0.67	1.64	0.73	1.57	0.71	1.61	0.70	1.61
4. Written Conventions	16	0.76	1.79	0.76	1.75	0.79	1.75	0.76	1.77
5. Writing Strategies	17	0.75	1.84	0.76	1.81	0.76	1.82	0.76	1.83
Grade 8									
1. Word Analysis and Vocabulary Development	9	0.66	1.34	0.61	1.34	0.70	1.31	0.61	1.35
2. Reading Comprehension	18	0.74	1.92	0.72	1.92	0.75	1.91	0.72	1.92
3. Literary Response and Analysis	15	0.70	1.77	0.70	1.77	0.72	1.75	0.68	1.77
4. Written Conventions	16	0.74	1.79	0.71	1.72	0.74	1.73	0.73	1.77
5. Writing Strategies	17	0.69	1.86	0.65	1.89	0.71	1.87	0.66	1.88
Grade 9									
1. Word Analysis and Vocabulary Development	8	0.61	1.17	0.57	1.09	0.59	1.14	0.60	1.14
2. Reading Comprehension	18	0.73	1.86	0.72	1.86	0.73	1.86	0.72	1.86
3. Literary Response and Analysis	16	0.61	1.79	0.61	1.74	0.61	1.76	0.62	1.78

Subscore Area	No. of Items	Male		Female		Not Econ. Dis.		Econ. Dis.	
		Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
4. Written Conventions	13	0.65	1.62	0.61	1.61	0.64	1.62	0.65	1.62
5. Writing Strategies	20	0.71	2.03	0.72	2.04	0.74	2.02	0.70	2.04
Grade 10									
1. Word Analysis and Vocabulary Development	8	0.69	1.17	0.66	1.12	0.64	1.09	0.68	1.17
2. Reading Comprehension	18	0.73	1.93	0.76	1.88	0.74	1.90	0.74	1.91
3. Literary Response and Analysis	16	0.66	1.79	0.71	1.75	0.62	1.75	0.69	1.78
4. Written Conventions	13	0.71	1.59	0.71	1.58	0.69	1.61	0.72	1.58
5. Writing Strategies	20	0.66	2.06	0.65	2.09	0.71	2.06	0.65	2.08
Grade 11									
1. Word Analysis and Vocabulary Development	8	0.56	1.31	0.37	1.23	0.59	1.24	0.49	1.29
2. Reading Comprehension	19	0.54	2.06	0.52	2.00	0.68	2.01	0.51	2.05
3. Literary Response and Analysis	17	0.71	1.87	0.79	1.74	0.81	1.74	0.78	1.81
4. Written Conventions	9	0.59	1.31	0.46	1.26	0.66	1.31	0.57	1.28
5. Writing Strategies	22	0.83	2.08	0.78	2.08	0.84	2.09	0.82	2.08

Table 8.B.11 Subscore Reliabilities and SEM for RLA by Special Services/Attendance in U.S. Schools

Subscore Area	No. of Items	No Spec. Serv.		Spec. Serv.		In U.S. Schools <12 Months		In U.S. Schools ≥12 Months	
		Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
Grade 2									
1. Word Analysis and Vocabulary Development	22	0.80	1.97	0.82	2.07	0.81	2.05	0.80	1.96
2. Reading Comprehension	15	0.77	1.69	0.78	1.71	0.78	1.71	0.77	1.69
3. Literary Response and Analysis	6	0.55	1.06	0.60	1.07	0.54	1.09	0.56	1.06
4. Written Conventions	14	0.77	1.64	0.78	1.65	0.73	1.68	0.77	1.64
5. Writing Strategies	8	0.52	1.29	0.44	1.26	0.50	1.26	0.52	1.29
Grade 3									
1. Word Analysis and Vocabulary Development	20	0.75	1.94	0.73	2.04	0.80	1.93	0.75	1.94
2. Reading Comprehension	15	0.72	1.76	0.66	1.74	0.72	1.75	0.72	1.75
3. Literary Response and Analysis	8	0.44	1.25	0.23	1.31	0.57	1.20	0.40	1.26
4. Written Conventions	13	0.68	1.64	0.60	1.66	0.70	1.61	0.66	1.64
5. Writing Strategies	9	0.55	1.38	0.43	1.34	0.54	1.37	0.56	1.38
Grade 4									
1. Word Analysis and Vocabulary Development	18	0.78	1.86	0.71	1.94	0.77	1.88	0.79	1.85
2. Reading Comprehension	15	0.73	1.75	0.65	1.74	0.70	1.74	0.76	1.74
3. Literary Response and Analysis	9	0.56	1.36	0.50	1.33	0.55	1.37	0.57	1.35
4. Written Conventions	18	0.80	1.83	0.72	1.91	0.81	1.84	0.80	1.82
5. Writing Strategies	15	0.68	1.76	0.55	1.62	0.70	1.75	0.67	1.75
Grade 5									
1. Word Analysis and Vocabulary Development	14	0.70	1.68	0.53	1.62	0.71	1.66	0.72	1.67
2. Reading Comprehension	16	0.67	1.83	0.40	1.78	0.69	1.81	0.66	1.83
3. Literary Response and Analysis	12	0.70	1.55	0.55	1.50	0.72	1.55	0.69	1.55
4. Written Conventions	17	0.77	1.83	0.75	1.80	0.81	1.81	0.74	1.83
5. Writing Strategies	16	0.71	1.81	0.58	1.77	0.74	1.78	0.70	1.83
Grade 6									
1. Word Analysis and Vocabulary Development	13	0.70	1.62	–	–	0.70	1.61	0.69	1.63
2. Reading Comprehension	17	0.69	1.91	–	–	0.67	1.92	0.71	1.90
3. Literary Response and Analysis	12	0.63	1.59	–	–	0.62	1.59	0.62	1.61
4. Written Conventions	16	0.75	1.80	–	–	0.74	1.80	0.77	1.78
5. Writing Strategies	17	0.71	1.87	–	–	0.74	1.85	0.66	1.89
Grade 7									
1. Word Analysis and Vocabulary Development	11	0.70	1.46	–	–	0.70	1.44	0.68	1.51
2. Reading Comprehension	18	0.74	1.90	–	–	0.73	1.91	0.78	1.90
3. Literary Response and Analysis	13	0.70	1.61	–	–	0.69	1.61	0.73	1.60
4. Written Conventions	16	0.76	1.77	–	–	0.77	1.75	0.75	1.79
5. Writing Strategies	17	0.76	1.83	–	–	0.76	1.83	0.78	1.81
Grade 8									
1. Word Analysis and Vocabulary Development	9	0.63	1.34	–	–	0.63	1.34	0.63	1.35
2. Reading Comprehension	18	0.73	1.92	–	–	0.75	1.90	0.66	1.96
3. Literary Response and Analysis	15	0.69	1.77	–	–	0.69	1.77	0.69	1.78
4. Written Conventions	16	0.74	1.75	–	–	0.72	1.74	0.76	1.78
5. Writing Strategies	17	0.68	1.88	–	–	0.70	1.87	0.59	1.90

Subscore Area	No. of Items	No Spec. Serv.		Spec. Serv.		In U.S. Schools <12 Months		In U.S. Schools ≥ 12 Months	
		Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
Grade 9									
1. Word Analysis and Vocabulary Development	8	0.59	1.14	–	–	0.59	1.13	0.62	1.19
2. Reading Comprehension	18	0.73	1.86	–	–	0.73	1.86	0.69	1.89
3. Literary Response and Analysis	16	0.61	1.77	–	–	0.62	1.76	0.58	1.83
4. Written Conventions	13	0.64	1.62	–	–	0.64	1.62	0.68	1.60
5. Writing Strategies	20	0.72	2.03	–	–	0.73	2.03	0.62	2.05
Grade 10									
1. Word Analysis and Vocabulary Development	8	0.68	1.15	–	–	0.65	1.13	0.70	1.19
2. Reading Comprehension	18	0.74	1.91	–	–	0.74	1.89	0.72	1.95
3. Literary Response and Analysis	16	0.68	1.78	–	–	0.66	1.76	0.70	1.80
4. Written Conventions	13	0.71	1.59	–	–	0.72	1.59	0.69	1.59
5. Writing Strategies	20	0.67	2.07	–	–	0.68	2.06	0.64	2.10
Grade 11									
1. Word Analysis and Vocabulary Development	8	0.51	1.28	–	–	0.45	1.27	0.63	1.29
2. Reading Comprehension	19	0.54	2.05	–	–	0.54	2.04	0.52	2.06
3. Literary Response and Analysis	17	0.78	1.81	–	–	0.79	1.79	0.77	1.86
4. Written Conventions	9	0.59	1.28	–	–	0.60	1.26	0.57	1.34
5. Writing Strategies	22	0.82	2.08	–	–	0.79	2.09	0.86	2.03

Table 8.B.12 Reliability of Classification for RLA, Grade Two

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
Decision Accuracy	«low1»-«high1»	0.03	0.03	0.00	0.00	0.00	0.06
	«low2»-«high2»	0.01	0.31	0.04	0.00	0.00	0.36
All-forms Average	«low3»-«high3»	0.00	0.04	0.27	0.03	0.00	0.34
	«low4»-«high4»	0.00	0.00	0.03	0.10	0.02	0.15
	«low5»-«high5»	0.00	0.00	0.00	0.03	0.06	0.09
Estimated Proportion Correctly Classified: Total = «special1», Proficient & Above = «special2»4							
Decision Consistency	0-16	0.03	0.03	0.00	0.00	0.00	0.06
	17-33	0.03	0.28	0.05	0.00	0.00	0.36
	34-47	0.00	0.05	0.24	0.04	0.00	0.34
Alternate Form	48-54	0.00	0.00	0.04	0.08	0.03	0.15
	55-65	0.00	0.00	0.00	0.03	0.06	0.09
Estimated Proportion Consistently Classified: Total = 0.69, Proficient & Above = 0.91							

Table 8.B.13 Reliability of Classification for RLA, Grade Three

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
Decision Accuracy	0-17	0.03	0.04	0.00	0.00	0.00	0.07
	18-30	0.01	0.30	0.05	0.00	0.00	0.36
	31-43	0.00	0.05	0.28	0.03	0.00	0.36
	44-52	0.00	0.00	0.04	0.11	0.01	0.16
All-forms Average	53-65	0.00	0.00	0.00	0.02	0.04	0.06
	Estimated Proportion Correctly Classified: Total = 0.76, Proficient & Above = 0.93						
Decision Consistency	0-17	0.03	0.04	0.00	0.00	0.00	0.07
	18-30	0.04	0.26	0.07	0.00	0.00	0.36
	31-43	0.00	0.07	0.24	0.04	0.00	0.36
Alternate Form	44-52	0.00	0.00	0.05	0.09	0.02	0.16
	53-65	0.00	0.00	0.00	0.02	0.04	0.06
Estimated Proportion Consistently Classified: Total = 0.66, Proficient & Above = 0.91							

Table 8.B.14 Reliability of Classification for RLA, Grade Four

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
Decision Accuracy	0-22	0.11	0.04	0.00	0.00	0.00	0.16
	23-35	0.02	0.22	0.04	0.00	0.00	0.28
	36-51	0.00	0.02	0.28	0.01	0.00	0.32
	52-60	0.00	0.01	0.04	0.10	0.02	0.17
All-forms Average	61-75	0.00	0.00	0.00	0.02	0.05	0.07
	Estimated Proportion Correctly Classified: Total = 0.77, Proficient & Above = 0.93						
Decision Consistency	0-22	0.10	0.04	0.01	0.00	0.00	0.16
	23-35	0.03	0.19	0.06	0.00	0.00	0.28
	36-51	0.00	0.04	0.25	0.03	0.00	0.32
Alternate Form	52-60	0.00	0.01	0.05	0.09	0.03	0.17
	61-75	0.00	0.00	0.00	0.02	0.05	0.07
Estimated Proportion Consistently Classified: Total = 0.68, Proficient & Above = 0.90							

Table 8.B.15 Reliability of Classification for RLA, Grade Five

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
Decision Accuracy	0-26	0.22	0.04	0.00	0.00	0.00	0.27
	27-33	0.03	0.10	0.04	0.00	0.00	0.17
	34-46	0.00	0.04	0.23	0.03	0.00	0.29
All-forms Average	47-57	0.00	0.00	0.03	0.12	0.01	0.16
	58-75	0.00	0.00	0.00	0.02	0.07	0.10
Estimated Proportion Correctly Classified: Total = 0.74, Proficient & Above = 0.93							
Decision Consistency	0-26	0.21	0.05	0.01	0.00	0.00	0.27
	27-33	0.05	0.08	0.05	0.00	0.00	0.17
	34-46	0.01	0.05	0.19	0.04	0.00	0.29
Alternate Form	47-57	0.00	0.00	0.04	0.10	0.02	0.16
	58-75	0.00	0.00	0.00	0.03	0.07	0.10
Estimated Proportion Consistently Classified: Total = 0.65, Proficient & Above = 0.91							

Table 8.B.16 Reliability of Classification for RLA, Grade Six

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
Decision Accuracy	0-23	0.16	0.05	0.00	0.00	0.00	0.21
	24-32	0.03	0.17	0.05	0.01	0.00	0.26
	33-45	0.00	0.03	0.22	0.01	0.00	0.26
All-forms Average	46-56	0.00	0.01	0.05	0.13	0.01	0.19
	57-75	0.00	0.00	0.00	0.02	0.05	0.08
Estimated Proportion Correctly Classified: Total = 0.73, Proficient & Above = 0.93							
Decision Consistency	0-23	0.15	0.06	0.01	0.00	0.00	0.21
	24-32	0.05	0.14	0.07	0.01	0.00	0.26
	33-45	0.00	0.04	0.19	0.03	0.00	0.26
Alternate Form	46-56	0.00	0.01	0.05	0.11	0.02	0.19
	57-75	0.00	0.00	0.00	0.02	0.05	0.07
Estimated Proportion Consistently Classified: Total = 0.63, Proficient & Above = 0.90							

Table 8.B.17 Reliability of Classification for RLA, Grade Seven

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
Decision Accuracy	0-23	0.11	0.03	0.00	0.00	0.00	0.15
	24-35	0.04	0.21	0.04	0.00	0.00	0.28
	36-48	0.00	0.04	0.20	0.03	0.00	0.27
All-forms Average	49-59	0.00	0.00	0.03	0.14	0.02	0.19
	60-75	0.00	0.00	0.00	0.02	0.08	0.11
Estimated Proportion Correctly Classified: Total = 0.75, Proficient & Above = 0.94							
Decision Consistency	0-23	0.10	0.05	0.00	0.00	0.00	0.15
	24-35	0.05	0.18	0.05	0.00	0.00	0.28
	36-48	0.00	0.05	0.18	0.04	0.00	0.27
Alternate Form	49-59	0.00	0.00	0.04	0.12	0.03	0.19
	60-75	0.00	0.00	0.00	0.03	0.08	0.11
Estimated Proportion Consistently Classified: Total = 0.65, Proficient & Above = 0.91							

Table 8.B.18 Reliability of Classification for RLA, Grade Eight

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
Decision Accuracy	0-20	0.05	0.05	0.00	0.00	0.00	0.10
	21-32	0.02	0.24	0.04	0.00	0.00	0.31
	33-46	0.00	0.04	0.26	0.03	0.00	0.33
	47-57	0.00	0.00	0.03	0.13	0.01	0.18
All-forms Average	58-75	0.00	0.00	0.00	0.02	0.06	0.09
	Estimated Proportion Correctly Classified: Total = 0.74, Proficient & Above = 0.93						
Decision Consistency	0-20	0.04	0.05	0.00	0.00	0.00	0.10
	21-32	0.05	0.21	0.06	0.00	0.00	0.31
	33-46	0.00	0.06	0.22	0.05	0.00	0.33
	47-57	0.00	0.00	0.04	0.11	0.02	0.18
Alternate Form	58-75	0.00	0.00	0.00	0.03	0.06	0.09
	Estimated Proportion Consistently Classified: Total = 0.64, Proficient & Above = 0.91						

Table 8.B.19 Reliability of Classification for RLA, Grade Nine

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
Decision Accuracy	0-22	0.05	0.03	0.00	0.00	0.00	0.08
	23-35	0.02	0.24	0.05	0.00	0.00	0.32
	36-48	0.00	0.04	0.28	0.04	0.00	0.36
	49-58	0.00	0.00	0.03	0.13	0.01	0.17
All-forms Average	59-75	0.00	0.00	0.00	0.03	0.04	0.07
	Estimated Proportion Correctly Classified: Total = 0.75, Proficient & Above = 0.92						
Decision Consistency	0-22	0.05	0.03	0.00	0.00	0.00	0.08
	23-35	0.04	0.21	0.07	0.00	0.00	0.32
	36-48	0.00	0.06	0.25	0.05	0.00	0.36
	49-58	0.00	0.00	0.04	0.11	0.02	0.17
Alternate Form	59-75	0.00	0.00	0.00	0.03	0.04	0.07
	Estimated Proportion Consistently Classified: Total = 0.65, Proficient & Above = 0.89						

Table 8.B.20 Reliability of Classification for RLA, Grade Ten

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
Decision Accuracy	0-22	0.07	0.03	0.00	0.00	0.00	0.10
	23-36	0.02	0.24	0.04	0.00	0.00	0.30
	37-48	0.00	0.04	0.24	0.04	0.01	0.33
	49-58	0.00	0.00	0.03	0.14	0.02	0.19
All-forms Average	59-75	0.00	0.00	0.00	0.02	0.06	0.08
	Estimated Proportion Correctly Classified: Total = 0.75, Proficient & Above = 0.92						
Decision Consistency	0-22	0.06	0.03	0.00	0.00	0.00	0.10
	23-36	0.03	0.21	0.06	0.00	0.00	0.30
	37-48	0.00	0.06	0.20	0.06	0.01	0.33
	49-58	0.00	0.00	0.04	0.12	0.03	0.19
Alternate Form	59-75	0.00	0.00	0.00	0.02	0.06	0.08
	Estimated Proportion Consistently Classified: Total = 0.65, Proficient & Above = 0.89						

Table 8.B.21 Reliability of Classification for RLA, Grade Eleven

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
Decision Accuracy	0-24	0.09	0.03	0.00	0.00	0.00	0.13
	25-38	0.02	0.25	0.05	0.00	0.00	0.31
	39-49	0.00	0.03	0.19	0.02	0.00	0.24
All-forms Average	50-58	0.00	0.01	0.05	0.13	0.02	0.21
	59-75	0.00	0.00	0.00	0.03	0.08	0.11
Estimated Proportion Correctly Classified: Total = 0.73, Proficient & Above = 0.91							
Decision Consistency	0-24	0.09	0.04	0.00	0.00	0.00	0.13
	25-38	0.03	0.22	0.06	0.00	0.00	0.31
	39-49	0.00	0.04	0.16	0.04	0.00	0.24
Alternate Form	50-58	0.00	0.01	0.06	0.11	0.03	0.21
	59-75	0.00	0.00	0.01	0.03	0.07	0.11
Estimated Proportion Consistently Classified: Total = 0.64, Proficient & Above = 0.88							

Appendix 8.C—Scaling and Equating Results

In 2014–15, administration of the STS for RLA was optional. For the majority of the grades, the number of students was lower than in previous administrations. To protect student privacy, frequency distribution is not shown if based on 10 or fewer student records. Because frequency distributions were not available for many raw score levels, no frequency distributions for any raw scores are included in Appendix 8.C.

Table 8.C.1 Conversion for RLA, Grade Two

Raw Scr.	Theta	Scale Score	Rprtd Score	Raw Scr.	Theta	Scale Score	Rprtd Score
0	N/A	150.0000	150	33	-0.4746	298.4645	298
1	-4.9648	150.0000	150	34	-0.4061	301.7395	302
2	-4.2412	150.0000	150	35	-0.3375	305.0207	305
3	-3.8057	150.0000	150	36	-0.2686	308.3138	308
4	-3.4882	154.3830	154	37	-0.1994	311.6250	312
5	-3.2357	166.4586	166	38	-0.1296	314.9605	315
6	-3.0242	176.5699	177	39	-0.0592	318.3270	318
7	-2.8411	185.3242	185	40	0.0120	321.7308	322
8	-2.6787	193.0855	193	41	0.0842	325.1814	325
9	-2.5323	200.0883	200	42	0.1575	328.6851	329
10	-2.3983	206.4946	206	43	0.2321	332.2513	332
11	-2.2743	212.4199	212	44	0.3082	335.8895	336
12	-2.1587	217.9499	218	45	0.3860	339.6125	340
13	-2.0499	223.1505	223	46	0.4659	343.4307	343
14	-1.9469	228.0718	228	47	0.5481	347.3589	347
15	-1.8490	232.7557	233	48	0.6329	351.4136	351
16	-1.7553	237.2353	237	49	0.7207	355.6137	356
17	-1.6653	241.5379	242	50	0.8121	359.9818	360
18	-1.5785	245.6866	246	51	0.9075	364.5445	365
19	-1.4946	249.7009	250	52	1.0077	369.3345	369
20	-1.4131	253.5976	254	53	1.1135	374.3917	374
21	-1.3337	257.3911	257	54	1.2259	379.7655	380
22	-1.2562	261.0945	261	55	1.3464	385.5259	386
23	-1.1805	264.7182	265	56	1.4766	391.7508	392
24	-1.1061	268.2732	268	57	1.6189	398.5568	399
25	-1.0330	271.7683	272	58	1.7768	406.1031	406
26	-0.9610	275.2118	275	59	1.9550	414.6262	415
27	-0.8899	278.6115	279	60	2.1612	424.4847	424
28	-0.8195	281.9746	282	61	2.4080	436.2850	436
29	-0.7498	285.3076	285	62	2.7191	451.1585	451
30	-0.6806	288.6180	289	63	3.1479	471.6588	472
31	-0.6117	291.9096	292	64	3.8640	505.8947	506
32	-0.5431	295.1898	295	65	N/A	600.0000	600

Note: Performance-level cut scores are highlighted.

Table 8.C.2 Conversion for RLA, Grade Three

Raw Scr.	Theta	Scale Score	Rprtd Score	Raw Scr.	Theta	Scale Score	Rprtd Score
0	N/A	150.0000	150	33	-0.2431	310.0341	310
1	-4.7666	150.0000	150	34	-0.1735	313.6641	314
2	-4.0423	150.0000	150	35	-0.1037	317.3003	317
3	-3.6060	150.0000	150	36	-0.0338	320.9491	321
4	-3.2877	151.2604	151	37	0.0366	324.6170	325
5	-3.0343	164.4751	164	38	0.1074	328.3108	328
6	-2.8220	175.5489	176	39	0.1789	332.0377	332
7	-2.6380	185.1446	185	40	0.2511	335.8053	336
8	-2.4747	193.6581	194	41	0.3243	339.6216	340
9	-2.3273	201.3466	201	42	0.3986	343.4952	343
10	-2.1923	208.3858	208	43	0.4741	347.4355	347
11	-2.0673	214.9015	215	44	0.5512	351.4563	351
12	-1.9506	220.9876	221	45	0.6300	355.5653	356
13	-1.8408	226.7143	227	46	0.7108	359.7775	360
14	-1.7368	232.1382	232	47	0.7938	364.1082	364
15	-1.6378	237.3037	237	48	0.8795	368.5750	369
16	-1.5430	242.2469	242	49	0.9682	373.1986	373
17	-1.4519	246.9977	247	50	1.0603	378.0033	378
18	-1.3640	251.5811	252	51	1.1565	383.0179	383
19	-1.2789	256.0182	256	52	1.2573	388.2786	388
20	-1.1963	260.3276	260	53	1.3637	393.8279	394
21	-1.1158	264.5242	265	54	1.4767	399.7209	400
22	-1.0372	268.6226	269	55	1.5977	406.0275	406
23	-0.9603	272.6348	273	56	1.7283	412.8394	413
24	-0.8848	276.5722	277	57	1.8710	420.2804	420
25	-0.8105	280.4437	280	58	2.0291	428.5278	429
26	-0.7373	284.2592	284	59	2.2076	437.8318	438
27	-0.6651	288.0268	288	60	2.4138	448.5880	449
28	-0.5936	291.7543	292	61	2.6606	461.4549	461
29	-0.5228	295.4490	295	62	2.9716	477.6726	478
30	-0.4524	299.1177	299	63	3.3999	500.0114	500
31	-0.3824	302.7672	303	64	4.1155	537.3282	537
32	-0.3127	306.4039	306	65	N/A	600.0000	600

Note: Performance-level cut scores are highlighted.

Table 8.C.3 Conversion for RLA, Grade Four

Raw Scr.	Theta	Scale Score	Rprtd Score	Raw Scr.	Theta	Scale Score	Rprtd Score
0	N/A	150.0000	150	38	-0.1572	305.9568	306
1	-4.7001	150.0000	150	39	-0.0985	309.1085	309
2	-3.9864	150.0000	150	40	-0.0398	312.2665	312
3	-3.5602	150.0000	150	41	0.0191	315.4342	315
4	-3.2517	150.0000	150	42	0.0784	318.6181	319
5	-3.0077	152.7382	153	43	0.1379	321.8204	322
6	-2.8044	163.6639	164	44	0.1980	325.0465	325
7	-2.6292	173.0803	173	45	0.2585	328.3012	328
8	-2.4746	181.3929	181	46	0.3197	331.5894	332
9	-2.3355	188.8657	189	47	0.3816	334.9164	335
10	-2.2088	195.6778	196	48	0.4443	338.2879	338
11	-2.0920	201.9577	202	49	0.5080	341.7119	342
12	-1.9833	207.8003	208	50	0.5728	345.1927	345
13	-1.8814	213.2784	213	51	0.6387	348.7387	349
14	-1.7852	218.4472	218	52	0.7061	352.3580	352
15	-1.6940	223.3524	223	53	0.7749	356.0598	356
16	-1.6069	228.0302	228	54	0.8455	359.8544	360
17	-1.5236	232.5105	233	55	0.9181	363.7532	364
18	-1.4434	236.8180	237	56	0.9928	367.7696	368
19	-1.3661	240.9737	241	57	1.0700	371.9187	372
20	-1.2913	244.9954	245	58	1.1500	376.2180	376
21	-1.2187	248.8992	249	59	1.2331	380.6871	381
22	-1.1480	252.6967	253	60	1.3200	385.3555	385
23	-1.0791	256.4011	256	61	1.4110	390.2472	390
24	-1.0117	260.0224	260	62	1.5068	395.4003	395
25	-0.9457	263.5698	264	63	1.6084	400.8592	401
26	-0.8810	267.0516	267	64	1.7167	406.6797	407
27	-0.8173	270.4761	270	65	1.8330	412.9335	413
28	-0.7545	273.8481	274	66	1.9592	419.7139	420
29	-0.6926	277.1754	277	67	2.0976	427.1560	427
30	-0.6315	280.4636	280	68	2.2515	435.4297	435
31	-0.5709	283.7182	284	69	2.4259	444.8005	445
32	-0.5109	286.9442	287	70	2.6281	455.6735	456
33	-0.4513	290.1464	290	71	2.8711	468.7310	469
34	-0.3921	293.3294	293	72	3.1783	485.2445	485
35	-0.3332	296.4977	296	73	3.6029	508.0707	508
36	-0.2744	299.6555	300	74	4.3150	546.3473	546
37	-0.2158	302.8072	303	75	N/A	600.0000	600

Note: Performance-level cut scores are highlighted.

Table 8.C.4 Conversion for RLA, Grade Five

Raw Scr.	Theta	Scale Score	Rprtd Score	Raw Scr.	Theta	Scale Score	Rprtd Score
0	N/A	150.0000	150	38	0.0104	315.5376	316
1	-4.4914	150.0000	150	39	0.0680	319.4014	319
2	-3.7796	150.0000	150	40	0.1257	323.2723	323
3	-3.3552	150.0000	150	41	0.1835	327.1555	327
4	-3.0484	150.0000	150	42	0.2416	331.0565	331
5	-2.8059	150.0000	150	43	0.3001	334.9806	335
6	-2.6041	150.0000	150	44	0.3590	338.9336	339
7	-2.4302	151.6895	152	45	0.4184	342.9216	343
8	-2.2768	161.9867	162	46	0.4784	346.9507	347
9	-2.1390	171.2388	171	47	0.5391	351.0276	351
10	-2.0134	179.6694	180	48	0.6007	355.1592	355
11	-1.8977	187.4383	187	49	0.6632	359.3566	359
12	-1.7901	194.6642	195	50	0.7267	363.6238	364
13	-1.6892	201.4355	201	51	0.7915	367.9717	368
14	-1.5941	207.8234	208	52	0.8576	372.4106	372
15	-1.5038	213.8834	214	53	0.9253	376.9520	377
16	-1.4177	219.6605	220	54	0.9946	381.6087	382
17	-1.3353	225.1917	225	55	1.0659	386.3951	386
18	-1.2562	230.5077	231	56	1.1394	391.3277	391
19	-1.1798	235.6352	236	57	1.2153	396.4246	396
20	-1.1059	240.5945	241	58	1.2941	401.7114	402
21	-1.0342	245.4064	245	59	1.3760	407.2102	407
22	-0.9645	250.0869	250	60	1.4615	412.9530	413
23	-0.8965	254.6505	255	61	1.5513	418.9770	419
24	-0.8301	259.1101	259	62	1.6459	425.3269	425
25	-0.7650	263.4770	263	63	1.7461	432.0583	432
26	-0.7012	267.7616	268	64	1.8531	439.2409	439
27	-0.6385	271.9732	272	65	1.9682	446.9634	447
28	-0.5767	276.1214	276	66	2.0931	455.3511	455
29	-0.5158	280.2111	280	67	2.2302	464.5547	465
30	-0.4556	284.2517	284	68	2.3828	474.8010	475
31	-0.3960	288.2496	288	69	2.5559	486.4171	486
32	-0.3370	292.2113	292	70	2.7568	499.9068	500
33	-0.2785	296.1425	296	71	2.9985	516.1307	516
34	-0.2203	300.0491	300	72	3.3044	536.6676	537
35	-0.1624	303.9366	304	73	3.7277	565.0820	565
36	-0.1047	307.8104	308	74	4.4385	600.0000	600
37	-0.0471	311.6758	312	75	N/A	600.0000	600

Note: Performance-level cut scores are highlighted.

Table 8.C.5 Conversion for RLA, Grade Six

Raw Scr.	Theta	Scale Score	Rprtd Score	Raw Scr.	Theta	Scale Score	Rprtd Score
0	N/A	150.0000	150	38	0.0047	319.8737	320
1	-4.4808	150.0000	150	39	0.0616	323.8729	324
2	-3.7682	150.0000	150	40	0.1186	327.8798	328
3	-3.3431	150.0000	150	41	0.1757	331.8997	332
4	-3.0359	150.0000	150	42	0.2332	335.9384	336
5	-2.7932	150.0000	150	43	0.2910	340.0016	340
6	-2.5912	150.0000	150	44	0.3492	344.0953	344
7	-2.4174	150.0000	150	45	0.4080	348.2257	348
8	-2.2642	160.3633	160	46	0.4673	352.3994	352
9	-2.1266	170.0359	170	47	0.5274	356.6233	357
10	-2.0013	178.8427	179	48	0.5883	360.9047	361
11	-1.8860	186.9519	187	49	0.6502	365.2553	365
12	-1.7788	194.4884	194	50	0.7131	369.6789	370
13	-1.6784	201.5456	202	51	0.7772	374.1873	374
14	-1.5838	208.1984	208	52	0.8427	378.7912	379
15	-1.4941	214.5054	215	53	0.9097	383.5025	384
16	-1.4086	220.5141	221	54	0.9785	388.3348	388
17	-1.3268	226.2637	226	55	1.0491	393.3030	393
18	-1.2483	231.7866	232	56	1.1220	398.4246	398
19	-1.1725	237.1108	237	57	1.1973	403.7186	404
20	-1.0993	242.2582	242	58	1.2754	409.2111	409
21	-1.0283	247.2503	247	59	1.3567	414.9261	415
22	-0.9593	252.1043	252	60	1.4416	420.8970	421
23	-0.8920	256.8354	257	61	1.5308	427.1624	427
24	-0.8262	261.4572	261	62	1.6247	433.7696	434
25	-0.7619	265.9819	266	63	1.7244	440.7765	441
26	-0.6987	270.4201	270	64	1.8308	448.2563	448
27	-0.6367	274.7818	275	65	1.9452	456.3020	456
28	-0.5756	279.0771	279	66	2.0696	465.0445	465
29	-0.5154	283.3112	283	67	2.2061	474.6420	475
30	-0.4559	287.4939	287	68	2.3582	485.3321	485
31	-0.3970	291.6321	292	69	2.5306	497.4573	497
32	-0.3387	295.7324	296	70	2.7310	511.5460	512
33	-0.2808	299.8011	300	71	2.9721	528.4974	528
34	-0.2233	303.8441	304	72	3.2775	549.9672	550
35	-0.1661	307.8673	308	73	3.7003	579.6889	580
36	-0.1091	311.8763	312	74	4.4107	600.0000	600
37	-0.0522	315.8766	316	75	N/A	600.0000	600

Note: Performance-level cut scores are highlighted.

Table 8.C.6 Conversion for RLA, Grade Seven

Raw Scr.	Theta	Scale Score	Rprtd Score	Raw Scr.	Theta	Scale Score	Rprtd Score
0	N/A	150.0000	150	38	-0.1957	310.0169	310
1	-4.7162	150.0000	150	39	-0.1373	313.6579	314
2	-4.0046	150.0000	150	40	-0.0788	317.3057	317
3	-3.5804	150.0000	150	41	-0.0201	320.9642	321
4	-3.2736	150.0000	150	42	0.0389	324.6413	325
5	-3.0311	150.0000	150	43	0.0982	328.3391	328
6	-2.8292	150.0000	150	44	0.1580	332.0638	332
7	-2.6552	156.6983	157	45	0.2182	335.8211	336
8	-2.5017	166.2702	166	46	0.2791	339.6167	340
9	-2.3637	174.8744	175	47	0.3407	343.4568	343
10	-2.2378	182.7187	183	48	0.4032	347.3478	347
11	-2.1218	189.9513	190	49	0.4665	351.2968	351
12	-2.0138	196.6816	197	50	0.5310	355.3150	355
13	-1.9126	202.9937	203	51	0.5966	359.4063	359
14	-1.8170	208.9508	209	52	0.6636	363.5819	364
15	-1.7263	214.6057	215	53	0.7321	367.8526	368
16	-1.6398	220.0000	220	54	0.8023	372.2302	372
17	-1.5569	225.1680	225	55	0.8745	376.7280	377
18	-1.4771	230.1380	230	56	0.9488	381.3616	381
19	-1.4002	234.9341	235	57	1.0256	386.1484	386
20	-1.3257	239.5767	240	58	1.1052	391.1089	391
21	-1.2534	244.0835	244	59	1.1879	396.2666	396
22	-1.1830	248.4699	248	60	1.2743	401.6526	402
23	-1.1144	252.7487	253	61	1.3649	407.2983	407
24	-1.0473	256.9322	257	62	1.4603	413.2466	413
25	-0.9815	261.0311	261	63	1.5614	419.5490	420
26	-0.9170	265.0545	265	64	1.6692	426.2705	426
27	-0.8535	269.0117	269	65	1.7851	433.4941	433
28	-0.7910	272.9090	273	66	1.9108	441.3284	441
29	-0.7293	276.7548	277	67	2.0487	449.9282	450
30	-0.6683	280.5555	281	68	2.2022	459.4933	459
31	-0.6080	284.3173	284	69	2.3760	470.3303	470
32	-0.5482	288.0459	288	70	2.5778	482.9096	483
33	-0.4888	291.7470	292	71	2.8201	498.0128	498
34	-0.4298	295.4257	295	72	3.1270	517.1415	517
35	-0.3710	299.0871	299	73	3.5512	543.5838	544
36	-0.3125	302.7363	303	74	4.2628	587.9419	588
37	-0.2541	306.3779	306	75	N/A	600.0000	600

Note: Performance-level cut scores are highlighted.

Table 8.C.7 Conversion for RLA, Grade Eight

Raw Scr.	Theta	Scale Score	Rprtd Score	Raw Scr.	Theta	Scale Score	Rprtd Score
0	N/A	150.0000	150	38	-0.1043	318.8006	319
1	-4.6982	150.0000	150	39	-0.0452	322.5068	323
2	-3.9826	150.0000	150	40	0.0138	326.2168	326
3	-3.5545	150.0000	150	41	0.0730	329.9362	330
4	-3.2441	150.0000	150	42	0.1325	333.6695	334
5	-2.9981	150.0000	150	43	0.1922	337.4219	337
6	-2.7928	150.0000	150	44	0.2524	341.1991	341
7	-2.6156	161.0704	161	45	0.3130	345.0064	345
8	-2.4590	170.9084	171	46	0.3742	348.8500	349
9	-2.3180	179.7636	180	47	0.4361	352.7360	353
10	-2.1893	187.8454	188	48	0.4987	356.6711	357
11	-2.0706	195.3032	195	49	0.5623	360.6645	361
12	-1.9600	202.2476	202	50	0.6269	364.7215	365
13	-1.8563	208.7635	209	51	0.6927	368.8518	369
14	-1.7583	214.9152	215	52	0.7597	373.0652	373
15	-1.6653	220.7559	221	53	0.8283	377.3723	377
16	-1.5766	226.3277	226	54	0.8986	381.7852	382
17	-1.4916	231.6652	232	55	0.9707	386.3173	386
18	-1.4099	236.7975	237	56	1.0451	390.9843	391
19	-1.3311	241.7489	242	57	1.1218	395.8038	396
20	-1.2548	246.5402	247	58	1.2013	400.7958	401
21	-1.1808	251.1897	251	59	1.2839	405.9875	406
22	-1.1088	255.7118	256	60	1.3702	411.4046	411
23	-1.0385	260.1213	260	61	1.4606	417.0827	417
24	-0.9699	264.4300	264	62	1.5558	423.0639	423
25	-0.9028	268.6486	269	63	1.6567	429.4004	429
26	-0.8369	272.7868	273	64	1.7643	436.1573	436
27	-0.7721	276.8534	277	65	1.8799	443.4183	443
28	-0.7084	280.8564	281	66	2.0053	451.2922	451
29	-0.6456	284.8037	285	67	2.1430	459.9396	460
30	-0.5835	288.7007	289	68	2.2961	469.5558	470
31	-0.5221	292.5549	293	69	2.4696	480.4521	480
32	-0.4614	296.3721	296	70	2.6710	493.1014	493
33	-0.4011	300.1579	300	71	2.9130	508.3031	508
34	-0.3412	303.9178	304	72	3.2193	527.5408	528
35	-0.2817	307.6572	308	73	3.6430	554.1516	554
36	-0.2224	311.3806	311	74	4.3542	598.8171	599
37	-0.1633	315.0935	315	75	N/A	600.0000	600

Note: Performance-level cut scores are highlighted.

Table 8.C.8 Conversion for RLA, Grade Nine

Raw Scr.	Theta	Scale Score	Rprtd Score	Raw Scr.	Theta	Scale Score	Rprtd Score
0	N/A	150.0000	150	38	-0.1456	308.3019	308
1	-4.7868	150.0000	150	39	-0.0846	312.0381	312
2	-4.0711	150.0000	150	40	-0.0235	315.7786	316
3	-3.6428	150.0000	150	41	0.0378	319.5309	320
4	-3.3321	150.0000	150	42	0.0993	323.2972	323
5	-3.0857	150.0000	150	43	0.1611	327.0836	327
6	-2.8799	150.0000	150	44	0.2234	330.8951	331
7	-2.7022	151.7657	152	45	0.2861	334.7372	335
8	-2.5449	161.3975	161	46	0.3495	338.6158	339
9	-2.4031	170.0757	170	47	0.4135	342.5368	343
10	-2.2736	178.0044	178	48	0.4783	346.5068	347
11	-2.1540	185.3292	185	49	0.5441	350.5325	351
12	-2.0425	192.1578	192	50	0.6109	354.6252	355
13	-1.9377	198.5726	199	51	0.6789	358.7886	359
14	-1.8387	204.6360	205	52	0.7483	363.0338	363
15	-1.7446	210.3998	210	53	0.8191	367.3712	367
16	-1.6546	215.9050	216	54	0.8916	371.8127	372
17	-1.5684	221.1851	221	55	0.9661	376.3714	376
18	-1.4854	226.2683	226	56	1.0427	381.0625	381
19	-1.4052	231.1781	231	57	1.1218	385.9031	386
20	-1.3275	235.9347	236	58	1.2036	390.9127	391
21	-1.2520	240.5564	241	59	1.2886	396.1177	396
22	-1.1786	245.0557	245	60	1.3772	401.5433	402
23	-1.1068	249.4479	249	61	1.4700	407.2241	407
24	-1.0366	253.7444	254	62	1.5676	413.2013	413
25	-0.9679	257.9554	258	63	1.6709	419.5255	420
26	-0.9003	262.0903	262	64	1.7809	426.2604	426
27	-0.8339	266.1575	266	65	1.8989	433.4874	433
28	-0.7685	270.1654	270	66	2.0267	441.3128	441
29	-0.7039	274.1195	274	67	2.1668	449.8921	450
30	-0.6401	278.0275	278	68	2.3224	459.4169	459
31	-0.5769	281.8955	282	69	2.4983	470.1902	470
32	-0.5143	285.7291	286	70	2.7022	482.6729	483
33	-0.4521	289.5339	290	71	2.9467	497.6447	498
34	-0.3904	293.3151	293	72	3.2555	516.5514	517
35	-0.3289	297.0776	297	73	3.6817	542.6461	543
36	-0.2677	300.8264	301	74	4.3954	586.3428	586
37	-0.2066	304.5663	305	75	N/A	600.0000	600

Note: Performance-level cut scores are highlighted.

Table 8.C.9 Conversion for RLA, Grade Ten

Raw Scr.	Theta	Scale Score	Rprtd Score	Raw Scr.	Theta	Scale Score	Rprtd Score
0	N/A	150.0000	150	38	-0.2102	304.7353	305
1	-4.8082	150.0000	150	39	-0.1503	308.6965	309
2	-4.0932	150.0000	150	40	-0.0904	312.6640	313
3	-3.6656	150.0000	150	41	-0.0303	316.6431	317
4	-3.3557	150.0000	150	42	0.0301	320.6391	321
5	-3.1101	150.0000	150	43	0.0908	324.6573	325
6	-2.9053	150.0000	150	44	0.1519	328.7035	329
7	-2.7285	150.0000	150	45	0.2135	332.7836	333
8	-2.5723	150.0000	150	46	0.2758	336.9037	337
9	-2.4316	157.6796	158	47	0.3387	341.0704	341
10	-2.3032	166.1786	166	48	0.4025	345.2907	345
11	-2.1847	174.0224	174	49	0.4671	349.5719	350
12	-2.0744	181.3279	181	50	0.5329	353.9257	354
13	-1.9708	188.1845	188	51	0.5998	358.3568	358
14	-1.8730	194.6595	195	52	0.6681	362.8769	363
15	-1.7801	200.8095	201	53	0.7379	367.4973	367
16	-1.6914	206.6787	207	54	0.8094	372.2306	372
17	-1.6064	212.3039	212	55	0.8828	377.0911	377
18	-1.5247	217.7154	218	56	0.9584	382.0951	382
19	-1.4458	222.9389	223	57	1.0365	387.2612	387
20	-1.3694	227.9963	228	58	1.1173	392.6112	393
21	-1.2952	232.9067	233	59	1.2013	398.1697	398
22	-1.2230	237.6868	238	60	1.2889	403.9703	404
23	-1.1526	242.3495	242	61	1.3807	410.0459	410
24	-1.0837	246.9089	247	62	1.4773	416.4420	416
25	-1.0162	251.3761	251	63	1.5796	423.2134	423
26	-0.9500	255.7612	256	64	1.6886	430.4291	430
27	-0.8848	260.0742	260	65	1.8056	438.1771	438
28	-0.8207	264.3215	264	66	1.9324	446.5727	447
29	-0.7574	268.5123	269	67	2.0715	455.7805	456
30	-0.6948	272.6538	273	68	2.2261	466.0122	466
31	-0.6329	276.7524	277	69	2.4010	477.5937	478
32	-0.5715	280.8144	281	70	2.6039	491.0242	491
33	-0.5106	284.8457	285	71	2.8472	507.1328	507
34	-0.4501	288.8521	289	72	3.1552	527.5176	528
35	-0.3899	292.8390	293	73	3.5804	555.6662	556
36	-0.3299	296.8117	297	74	4.2930	600.0000	600
37	-0.2700	300.7754	301	75	N/A	600.0000	600

Note: Performance-level cut scores are highlighted.

Table 8.C.10 Conversion for RLA, Grade Eleven

Raw Scr.	Theta	Scale Score	Rprtd Score	Raw Scr.	Theta	Scale Score	Rprtd Score
0	N/A	150.0000	150	38	-0.1851	296.0833	296
1	-4.7579	150.0000	150	39	-0.1260	300.5114	301
2	-4.0431	150.0000	150	40	-0.0667	304.9472	305
3	-3.6159	150.0000	150	41	-0.0072	309.3966	309
4	-3.3064	150.0000	150	42	0.0525	313.8658	314
5	-3.0614	150.0000	150	43	0.1125	318.3607	318
6	-2.8571	150.0000	150	44	0.1730	322.8878	323
7	-2.6809	150.0000	150	45	0.2340	327.4538	327
8	-2.5252	150.0000	150	46	0.2957	332.0657	332
9	-2.3852	150.0000	150	47	0.3580	336.7309	337
10	-2.2575	150.0000	150	48	0.4212	341.4572	341
11	-2.1398	150.0000	150	49	0.4852	346.2532	346
12	-2.0301	158.0060	158	50	0.5504	351.1323	351
13	-1.9273	165.7022	166	51	0.6168	356.0992	356
14	-1.8302	172.9664	173	52	0.6845	361.1676	361
15	-1.7381	179.8624	180	53	0.7538	366.3502	366
16	-1.6502	186.4407	186	54	0.8247	371.6613	372
17	-1.5660	192.7426	193	55	0.8976	377.1171	377
18	-1.4850	198.8028	199	56	0.9727	382.7360	383
19	-1.4069	204.6505	205	57	1.0503	388.5392	389
20	-1.3312	210.3104	210	58	1.1306	394.5513	395
21	-1.2578	215.8041	216	59	1.2141	400.7998	401
22	-1.1864	221.1504	221	60	1.3013	407.3242	407
23	-1.1167	226.3648	226	61	1.3926	414.1602	414
24	-1.0486	231.4625	231	62	1.4888	421.3599	421
25	-0.9819	236.4562	236	63	1.5907	428.9853	429
26	-0.9164	241.3574	241	64	1.6993	437.1144	437
27	-0.8520	246.1775	246	65	1.8160	445.8469	446
28	-0.7886	250.9237	251	66	1.9425	455.3131	455
29	-0.7260	255.6064	256	67	2.0813	465.6999	466
30	-0.6642	260.2338	260	68	2.2356	477.2459	477
31	-0.6030	264.8132	265	69	2.4103	490.3199	490
32	-0.5423	269.3517	269	70	2.6129	505.4865	505
33	-0.4821	273.8560	274	71	2.8560	523.6822	524
34	-0.4223	278.3326	278	72	3.1638	546.7169	547
35	-0.3628	282.7877	283	73	3.5889	578.5299	579
36	-0.3035	287.2273	287	74	4.3014	600.0000	600
37	-0.2443	291.6572	292	75	N/A	600.0000	600

Note: Performance-level cut scores are highlighted

Chapter 9: Quality Control Procedures

Rigorous quality control procedures were implemented throughout the test development, administration, scoring, and reporting processes. As part of this effort, Educational Testing Service (ETS) maintains an Office of Testing Integrity (OTI) that resides in the ETS legal department. The OTI provides quality assurance services for all testing programs administered by ETS. In addition, the Office of Professional Standards Compliance at ETS publishes and maintains the *ETS Standards for Quality and Fairness*, which supports the OTI's goals and activities. The purposes of the *ETS Standards for Quality and Fairness* are to help ETS design, develop, and deliver technically sound, fair, and useful products and services; and to help the public and auditors evaluate those products and services.

In addition, each department at ETS that is involved in the testing cycle designs and implements an independent set of procedures to ensure the quality of its products. In the next sections, these procedures are described.

Quality Control of Item Development

The item development process for the Standards-based Tests in Spanish (STS) for Reading/Language Arts (RLA) prior to the 2014–15 administration is described in detail in Chapter 3, starting on page 37; there was no new item development in 2014–15 because the forms were reused. The next sections highlight elements of the process devoted specifically to the quality control of the items that were previously developed and reused during the 2014–15 STS for RLA administration.

Item Specifications

ETS maintained item specifications for each STS for RLA and developed an item utilization plan to guide the development of the items for each content area. Item writing emphasis was determined in consultation with the California Department of Education (CDE). Adherence to the specifications ensured the maintenance of quality and consistency in the item development process.

Item Writers

The items for each STS for RLA were written by item writers with a thorough understanding of the California content standards. The item writers were carefully screened and selected by senior ETS content staff and approved by the CDE. Only those with strong content and teaching backgrounds were invited to participate in an extensive training program for item writers.

Internal Contractor Reviews

Once items were written, ETS assessment specialists made sure that each item underwent an intensive internal review process. Every step of this process is designed to produce items that exceed industry standards for quality. For the STS for RLA, it included three rounds of content reviews, two rounds of editorial reviews, an internal fairness review, and a high-level review and approval by a content-area director. A carefully designed and monitored workflow and detailed checklists helped to ensure that all items met the specifications for the process.

Content Review

ETS assessment specialists made sure that the test items and related materials complied with ETS's written guidelines for clarity, style, accuracy, and appropriateness, and with approved item specifications.

The artwork and graphics for the items were created during the internal content review period so assessment specialists could evaluate the correctness and appropriateness of the art early in the item development process. ETS selected visuals that were relevant to the item content and that were easily understood so students would not struggle to determine the purpose or meaning of the questions.

Editorial Review

Another step in the ETS internal review process involved a team of specially trained editors who checked questions for clarity, correctness of language, grade-level appropriateness of language, adherence to style guidelines, and conformity to acceptable item-writing practices. The editorial review also included rounds of copyediting and proofreading. ETS strives for error-free items beginning with the initial rounds of review.

Fairness Review

One of the final steps in the ETS internal review process is to have all items and stimuli reviewed for fairness. Only ETS staff members who had participated in the ETS Fairness Training, a rigorous internal training course, conducted this bias and sensitivity review. These staff members had been trained to identify and eliminate test questions that contained content that could be construed as offensive to, or biased against, members of specific ethnic, racial, or gender groups.

Assessment Director Review

As a final quality control step, the content area's assessment director or another senior-level content reviewer read each item before it was presented to the CDE.

Assessment Review Panel Review

The Assessment Review Panels (ARPs) were committees that advised the CDE and ETS on areas related to item development for the STS for RLA. The ARPs were responsible for reviewing all newly developed items for alignment to the California content standards. The ARPs also reviewed the items for accuracy of content, clarity of phrasing, and quality. See page 41 in Chapter 3 for additional information on the function of ARPs within the item-review process.

Statewide Pupil Assessment Review Panel Review

The Statewide Pupil Assessment Review (SPAR) panel was responsible for reviewing and approving the achievement tests that were used statewide for the testing of students in California public schools in grades two through eleven. The SPAR panel representatives ensured that the test items conformed to the requirements of *Education Code* Section 60602. If the SPAR panel rejected specific items, the items were replaced with other items. See page 43 in Chapter 3 for additional information on the function of the SPAR panel within the item-review process.

Data Review of Field-tested Items

ETS field-tested newly developed items to obtain statistical information about item performance. This information was used to evaluate items that were candidates for use in operational test forms. These items that were flagged after field-test and operational use were examined carefully at data review meetings, where content experts discussed items that had poor statistics and did not meet the psychometric criteria for item quality. The CDE defined the criteria for acceptable or unacceptable item statistics. These criteria ensured that the item (1) had an appropriate level of difficulty for the target population; (2) discriminated well between examinees that differ in ability; and (3) conformed well to the statistical model underlying the measurement of the intended constructs. The results of

analyses for differential item functioning (DIF) were used to make judgments about the appropriateness of items for various subgroups when the items were first used.

The ETS content experts made recommendations about whether to accept or reject each item for inclusion in the California item bank. The CDE content experts reviewed the recommendations and made the final decision on each item.

The field-test items that appeared in the STS for RLA administered in 2014–15 were statistically reviewed in data review meetings the year they were originally administered. There was no data review of field-test items in 2014–15. See Table 8.4 on page 111 for the list of the original administrations of each test administered in 2014–15.

Quality Control of the Item Bank

After the data review, items were placed in the item bank along with their statistics and reviewers' evaluations of their quality. ETS then delivered the items to the CDE through the California electronic item bank. The item bank database is maintained by a staff of application systems programmers, led by the Item Bank Manager, at ETS. All processes are logged, all change requests—including item bank updates for item availability status—are tracked, and all output and California item bank deliveries are quality-controlled for accuracy.

Quality of the item bank and secure transfer of the California item bank to the CDE are very important. The ETS internal item bank database resides on a server within the ETS firewall; access to the SQL Server database is strictly controlled by means of system administration. The electronic item banking application includes a login/password system to authorize access to the database or designated portions of the database. In addition, only users authorized to access the specific database are able to use the item bank. Users are authorized by a designated administrator at the CDE and at ETS.

ETS has extensive experience in accurate and secure data transfer of many types, including CDs, secure remote hosting, secure Web access, and secure file transfer protocol (SFTP), which is the current method used to deliver the California electronic item bank to the CDE. In addition, all files posted on the SFTP site by the item bank staff are encrypted with a password.

The measures taken for ensuring the accuracy, confidentiality, and security of electronic files are as follows:

- Electronic forms of test content, documentation, and item banks are backed up electronically, with the backup media kept off site, to prevent loss from system breakdown or a natural disaster.
- The offsite backup files are kept in secure storage, with access limited to authorized personnel only.
- Advanced network security measures are used to prevent unauthorized electronic access to the item bank.

Quality Control of Test Form Development

The ETS Assessment Development group is committed to providing the highest quality product to the students of California and has in place a number of quality control (QC) checks to ensure that outcome. During the item development process, there were multiple senior reviews of items and passages, including one by the assessment director. Test forms certification was a formal quality control process established as a final checkpoint prior to

printing. In it, content, editorial, and senior development staff reviewed test forms for accuracy and clueing issues.

ETS also included quality checks throughout preparation of the form planners. A form planner specifications document was developed by the test development team lead with input from ETS's item bank and statistics groups; this document was then reviewed by all team members who built forms at a training session specific to form planners before the form-building process started. After trained content team members signed off on a form planner, a representative from the internal QC group reviewed each file for accuracy against the specifications document. Assessment directors reviewed and signed off on form planners prior to processing.

As processes are refined and enhanced, ETS implements further QC checks as appropriate.

Quality Control of Test Materials

Collecting Test Materials

Once the tests are administered, local educational agencies (LEAs) return scorable and nonscorable materials within five working days after the last selected testing day of each test administration period. The freight return kits provided to the LEAs contain color-coded labels identifying scorable and nonscorable materials and labels with bar-coded information identifying the school and district. The LEAs apply the appropriate labels and number the cartons prior to returning the materials to the processing center by means of their assigned carrier. The use of the color-coded labels streamlines the return process.

All scorable and nonscorable materials are delivered to the ETS scanning and scoring facilities in Ewing, New Jersey. ETS closely monitor the return of materials. The California Technical Assistance Center (CaITAC) at ETS monitors returns and notifies LEAs that do not return their materials in a timely manner. CaITAC contacts the LEA California Assessment of Student Performance and Progress (CAASPP) coordinators and works with them to facilitate the return of the test materials.

Processing Test Materials

Upon receipt of the test materials, ETS uses precise inventory and test processing systems, in addition to quality assurance procedures, to maintain an up-to-date accounting of all the testing materials within its facilities. The materials are removed carefully from the shipping cartons and examined for a number of conditions, including physical damage, shipping errors, and omissions. A visual inspection to compare the number of students recorded on the School and Grade Identification (SGID) sheets with the number of answer documents in the stack is also conducted.

ETS's image scanning process captures security information electronically and compares scorable material quantities reported on the SGIDs to actual documents scanned. LEAs are contacted by phone if there are any missing shipments or the quantity of materials returned appears to be less than expected.

Quality Control of Scanning

Before any CAASPP documents are scanned, ETS conducts a complete check of the scanning system. ETS creates test decks for every test and form. Each test deck consists of approximately 700 answer documents marked to cover response ranges, demographic data, blanks, double marks, and other responses. Fictitious students are created to verify

that each marking possibility is processed correctly by the scanning program. The output file generated as a result of this activity is thoroughly checked against each answer document after every stage to verify that the scanner is capturing marks correctly. When the program output is confirmed to match the expected results, a scan program release form is signed and the scan program is placed in the production environment under configuration management.

The intensity levels of each scanner are constantly monitored for quality control purposes. Intensity diagnostics sheets are run before and during each batch to verify that the scanner is working properly. In the event that a scanner fails to properly pick up items on the diagnostic sheets, the scanner is recalibrated to work properly before being allowed to continue processing student documents.

Documents received in poor condition (torn, folded, or water-stained) that could not be fed through the high-speed scanners are either scanned using a flat-bed scanner or keyed into the system manually.

Quality Control of Image Editing

Prior to submitting any CAASPP operational documents through the image editing process, ETS creates a mock set of documents to test all of the errors listed in the edit specifications. The set of test documents is used to verify that each image of the document is saved so that an editor would be able to review the documents through an interactive interface. The edits are confirmed to show the appropriate error, the correct image to edit the item, and the appropriate problem and resolution text that instructs the editor on the actions that should be taken.

Once the set of mock test documents is created, the image edit system completes the following procedures:

1. Scan the set of test documents.
2. Verify that the images from the documents are saved correctly.
3. Verify that the appropriate problem and resolution text displays for each type of error.
4. Submit the post-edit program to assure that all errors have been corrected.

ETS checks the post file against expected results to ensure the appropriate corrections are made. The post file will have all keyed corrections and any defaults from the edit specifications.

Quality Control of Answer Document Processing and Scoring

Accountability of Answer Documents

In addition to the quality control checks carried out in scanning and image editing, the following manual quality checks are conducted to verify that the answer documents are correctly attributed to the students, schools, LEAs, and subgroups, and document counts are compared to the SGIDs.

Any discrepancies identified in the steps outlined above are followed up by ETS staff with the LEAs for resolution.

Processing of Answer Documents

Prior to processing operational answer documents and executing subsequent data processing programs, ETS conducts an end-to-end test. As part of this test, ETS prepares approximately 700 test cases covering all tests and many scenarios designed to exercise

particular business rule logic. ETS marks answer documents for those 700 test cases. They are then scanned, scored, and aggregated. The results at various inspection points are checked by psychometricians and Data Quality Services staff. Additionally, a post-scan test file of approximately 50,000 records across the CAASPP System is scored and aggregated to test a broader range of scoring and aggregation scenarios. These procedures assure that students and LEAs receive the correct scores when the actual scoring process is carried out. In 2014–15, end-to-end testing also included the inspection of results in electronic reporting.

Scoring and Reporting Specifications

ETS develops standardized scoring procedures and specifications so testing materials are processed and scored accurately. These documents include the Scoring Rules specifications and the Include Indicators specifications. Each is explained in detail in Chapter 7, starting on page 81. The scoring specifications are reviewed and revised by the CDE and ETS each year. After a version that all parties endorse is finalized, the CDE issues a formal approval of the scoring and reporting specifications.

Storing Answer Documents

After the answer documents have been scanned, edited, and scored, and have cleared the clean-post process, they are palletized and placed in the secure storage facilities at ETS. The materials are stored until October 31 of each year, after which ETS requests permission to destroy the materials. After receiving CDE approval, the materials are destroyed in a secure manner.

Quality Control of Psychometric Processes

Score Key Verification Procedures

ETS takes various necessary measures to ascertain that the scoring keys are applied to the student responses as expected and the student scores are computed accurately. Scoring keys, provided in the form planners, are produced by ETS and verified thoroughly by performing multiple quality control checks. The form planners contain the information about an assembled test form; other information in the form planner includes the test name, administration year, subscore identification, and standards and statistics associated with each item. The quality control checks that are performed before keys are finalized are listed on page 82 in Chapter 7.

Quality Control of Item Analyses and the Equating Process

When the forms were first administered, the psychometric analyses conducted at ETS underwent comprehensive quality checks by a team of psychometricians and data analysts. Detailed checklists were consulted by members of the team for each of the statistical procedures performed on each STS for RLA following its original administration. Quality assurance checks also included a comparison of the current year's statistics to statistics from previous years. The results of preliminary classical item analyses that provided a check on scoring keys were also reviewed by a senior psychometrician. The items that were flagged for questionable statistical attributes were sent to test development staff for their review; their comments were reviewed by the psychometricians before items were approved to be included in the equating process.

The results of the equating process were reviewed by a psychometric manager in addition to the aforementioned team of psychometricians and data analysts. If the senior psychometrician and the manager reached a consensus that an equating result did not

conform to the norm, special binders were prepared for review by senior psychometric advisors at ETS, along with several pieces of informative analyses to facilitate the process.

When the forms were equated following their original administration, a few additional checks were performed for the calibration, scaling, and scoring table creation processes, as described below.

Calibrations

During the calibration that was conducted for the original administration of each form and that is described in more detail in Chapter 2 starting on page 14, checks were made to ascertain that the correct options for the analyses are selected. Checks were also made on the number of items, number of examinees with valid scores, item response theory (IRT) Rasch item difficulty estimates, standard errors for the Rasch item difficulty estimates, and the match of selected statistics to the results on the same statistics obtained during preliminary item analyses. Psychometricians also performed detailed reviews of plots and statistics to investigate if the model fit the data.

Scaling

During the scaling that was conducted for the original administration of each form, checks were made to ensure the following:

- The correct items are used for linking;
- The scaling evaluation process, including stability analysis and subsequent removal of items from the linking set (if any), was implemented according to specification (see details in the “Evaluation of Scaling” section in Chapter 8 of the original year’s technical report); and
- The resulting scaling constants were correctly applied to transform the new item difficulty estimates onto the item bank scale.

Scoring Tables

Once the equating activities were complete and raw-score-to-scale score conversion tables were generated after the original administration of each content-area test, the psychometricians carried out quality control checks on each scoring table. Scoring tables were checked to verify the following:

- All raw scores were included in the tables;
- Scale scores increased as raw scores increased;
- The minimum reported scale score was 150 and maximum reported scale score was 600; and
- The cut points for the performance levels were correctly identified.

As a check on the reasonableness of the performance levels, when the tests were originally administered, psychometricians compared results from the current year with results from the past year at the cut points and the percentage of students in each performance level within the equating samples. After all quality control steps were completed and any differences were resolved, a senior psychometrician inspected the scoring tables as the final step in quality control.

During the current administration, the data derived from previous item analyses are used to pre-equate the 2014–15 results. Key checks and classical item analyses as well as associated quality assurance checks are also conducted on the current data.

In addition, the scoring tables are reused and are checked against the scoring tables in the reuse-year technical report to ensure exact match. In addition, prior to reporting in 2014–15, every regular and special-version multiple-choice test was certified by ETS prior to being included in electronic reporting. To certify a test, psychometricians gathered a certain number of test cases and verified the accurate application of scoring keys and conversion tables.

Score Verification Process

ETS utilizes the raw-to-scale scoring tables to assign scale scores for each student and verifies scale scores by independently generating the scale scores for students in a small number of LEAs and comparing these scores. The selection of LEAs is based on the availability of data for all schools included in those LEAs, known as “pilot LEAs.”

Year-to-Year Comparison Analyses

Year-to-year comparison analyses are conducted each year for quality control of the scoring procedure in general and as reasonableness checks for the STS for RLA results. The year-to-year comparison analyses uses over 90 percent of the entire testing population to look at the tendencies and trends for the state as a whole as well as a few large LEAs.

The results of the year-to-year comparison analyses are provided to the CDE, and their reasonableness is jointly discussed. Any anomalies in the results are investigated further, and scores are released only after explanations that satisfy both the CDE and ETS are obtained.

Offloads to Test Development

During the original administration of the STS for RLA forms that are reused in 2014–15, the statistics based on classical item analyses and the IRT analyses were obtained at two different times in the testing cycle. The first time, the statistics were obtained on the equating samples to ensure the quality of equating and then on larger sample sizes to ensure the stability of the statistics that were to be used for future test assembly. The resulting statistics for all items were provided to test development staff in specially designed Excel spreadsheets called “statistical offloads.” The offloads were thoroughly checked by the psychometric staff before their release for test development review.

During the 2014–15 administration, only statistics based on classical item analyses of the operational items are included in the statistical offloads.

Quality Control of Reporting

For the quality control of various CAASPP student and summary reports, the following four general areas are evaluated:

1. Comparing report formats to input sources from the CDE-approved samples
2. Validating and verifying the report data by querying the appropriate student data
3. Evaluating the production print execution performance by comparing the number of report copies, sequence of report order, and offset characteristics to the CDE’s requirements
4. Proofreading reports by the CDE and ETS prior to any LEA mailings

All reports are required to include a single, accurate county/district/school (CDS) code, a charter school number (if applicable), an LEA name, and a school name. All elements conform to the CDE’s official CDS code and naming records. From the start of processing through scoring and reporting, the CDS Master File is used to verify and confirm accurate

codes and names. The CDS Master File is provided by the CDE to ETS throughout the year as updates are available.

After the reports are validated against the CDE's requirements, a set of reports for pilot LEAs is provided to the CDE and ETS for review and approval. ETS sends paper reports on the actual report forms, foldered as they are expected to look in production. The CDE and ETS review and sign off on the report package after a thorough review.

Upon the CDE's approval of the reports generated from the pilot LEAs, ETS proceeds with the first production batch test. The first production batch is selected to validate a subset of LEAs that contains examples of key reporting characteristics representative of the state as a whole. The first production batch test incorporates CDE-selected LEAs and provides the last check prior to generating all reports and mailing them to the LEAs.

Electronic Reporting

Because results were pre-equated, students' scale scores and performance levels for STS for RLA multiple-choice tests were made available to LEAs prior to the printing of paper reports. The reporting module in the Test Operations Management System made it possible for LEAs to securely download an electronic reporting file containing these results.

Before an LEA could download a student data file, ETS statisticians approved a QC file of test results data and ETS IT successfully processed it. Once the data were deemed reliable and ETS processed a scorable answer document for every student who took an STS for RLA in that test administration for the LEA, the LEA was notified that these results were available.

Excluding Student Scores from Summary Reports

ETS provides specifications to the CDE that document when to exclude student scores from summary reports. These specifications include the logic for handling answer documents that, for example, indicate the student tested but marked no answers, was absent, was not tested due to parent/guardian request, or did not complete the test due to illness. The methods for handling other anomalies are also covered in the specifications.

Reference

Educational Testing Service. (2002). *ETS standards for quality and fairness*. Princeton, NJ: Author.

Chapter 10: Historical Comparisons

Base Year Comparisons

Historical comparisons of the Standards-based Tests in Spanish (STS) for Reading/Language Arts (RLA) results are routinely performed to identify the trends in examinee performance and test characteristics over time. Such comparisons were performed for RLA in grades two through eleven over a period of the three most recent years of administration—2013, 2014, and 2015—and the base year.

The indicators of examinee performance include the mean and standard deviation of scale scores, observed score ranges, and the percentage of examinees classified into proficient and advanced performance levels. Test characteristics are compared by looking at the mean proportion correct, overall reliability and standard errors of measurement (SEM), as well as the mean item response theory (IRT) *b*-value for each STS for RLA.

The base year of each STS for RLA refers to the year in which the base score scale was established. Operational forms administered in the years following the base year are linked to the base year score scale using procedures described in Chapter 2.

The base years for the STS for RLA are presented in Table 10.1.

Table 10.1 Base Years for the STS for RLA

STS	Base Year
2 RLA	2009
3 RLA	2009
4 RLA	2009
5 RLA	2010
6 RLA	2010
7 RLA	2010
8 RLA	2013
9 RLA	2013
10 RLA	2013
11 RLA	2013

The base years differ over STS for RLA by grade. Reasons for these differences are as follows:

- The STS for RLA in grades two through four were first administered operationally in spring 2007. Percent-correct scores were reported in the 2007 and 2008 administrations. A standard setting was held in fall 2008 to establish new cut scores for the below basic, basic, proficient, and advanced performance levels. Spring 2009 was the first administration in which test results were reported using the new scales and cut scores for the four performance levels (below basic, basic, proficient, and advanced); thus, 2009 became the base year for these tests.
- The STS for RLA in grades five through seven were first administered operationally in spring 2008. Percent-correct scores were reported in the 2008 and 2009 administrations. A standard setting was held in fall 2009 to establish new cut scores for the below basic, basic, proficient, and advanced performance levels. Spring 2010 was the first administration in which test results were reported using the new scales and cut scores for the four performance levels (below basic, basic, proficient, and advanced); thus, 2010 became the base year for these tests.

- The STS for RLA in grades eight through eleven were first administered operationally in spring 2009. Percent-correct scores were reported in the 2009 through 2012 administrations. The performance standards for those tests were developed in November 2011 and adopted by the State Board of Education in July 2012. Spring 2013 was the first administration in which test results were reported using the new scales and cut scores for the four performance levels (below basic, basic, proficient, and advanced); thus, 2013 became the base year for these tests.

Examinee Performance

Table 10.A.1 on page 160 contains the number of examinees assessed and the means and standard deviations of examinees' scale scores in the base year and subsequent administrations for each grade-level STS for RLA in grades two through eleven for the base year and in 2013, 2014, and 2015. As noted in previous chapters, the STS for RLA reporting scales range from 150 to 600 for all of the tests.

STS for RLA scale scores are used to classify student results into one of five performance levels: far below basic, below basic, basic, proficient, and advanced. The percentages of students qualifying for the proficient and advanced levels are presented in Table 10.A.2 and Table 10.A.3 on page 160; please note that this information may differ slightly from information found on the CDE California Assessment of Student Performance and Progress (CAASPP) reporting Web page at <http://caaspp.cde.ca.gov> due to differing dates on which data were accessed. The goal is for all students to achieve at or above the proficient level by 2014.

Table 10.A.4 through Table 10.A.6 show the distribution of scale scores observed in the base year, which differs according to test, and subsequent administrations in 2013, 2014, and 2015 for the tests in grades two through eleven. Frequency counts are provided for each scale score interval of 30. A frequency count of "N/A" indicates that there are no obtainable scale scores within that scale-score range. For all STS for RLA, a minimum score of 300 is required for a student to reach the basic level of performance, and a minimum score of 350 is required for a student to reach the proficient level of performance.

Test Characteristics

The item and test analysis results of the STS for RLA over the comparison years indicate that the STS for RLA meets the technical criteria established in professional standards for high-stakes tests. In addition, every year, efforts were made to improve the technical quality of each STS for RLA.

Table 10.B.1 in Appendix 10.B, which starts on page 162, presents the average proportion-correct values for the operational items in each STS for RLA based on the equating samples. The mean proportion correct is affected by both the difficulty of the items and the abilities of the students administered the items.

Table 10.B.2 shows the mean equated IRT *b*-values for the STS for RLA operational items based on the equating samples. The mean equated IRT *b*-values reflect only average item difficulty. Please note that comparisons of mean *b*-values should be made only within a given test; they should not be compared across grade-level tests.

The average point-biserial correlations for all of the STS for RLA are presented in Table 10.B.3. The reliabilities and standard error of measurement (SEM) expressed in raw score units appear in Table 10.B.4 and Table 10.B.5. Like the average proportion correct, point-biserial correlations and reliabilities are affected by both item characteristics and student characteristics.

Appendix 10.A—Historical Comparisons Tables, Examinee Performance

Table 10.A.1 Number of Examinees Tested, Scale Score Means, and Standard Deviations of STS for RLA Across Base Year, 2013, 2014, and 2015

STS	Number of Examinees (valid scores)				Scale Score Mean and Standard Deviation							
	Base	2013	2014	2015	Base		2013		2014		2015	
					Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
2 RLA	13,300	8,854	1,930	1,879	333	53	329	55	319	55	313	49
3 RLA	8,992	6,259	1,388	1,643	333	51	333	53	312	45	314	47
4 RLA	4,897	3,903	803	934	325	51	326	54	316	52	310	51
5 RLA	3,049	2,743	645	730	318	56	317	59	313	59	313	61
6 RLA	1,962	1,725	482	440	318	54	323	62	318	62	311	59
7 RLA	1,587	1,414	460	455	322	52	333	60	323	57	318	60
8 RLA	1,083	1,083	439	457	330	58	330	58	323	57	317	53
9 RLA	1,980	1,980	956	810	324	54	324	54	320	54	315	50
10 RLA	1,177	1,177	388	285	327	55	327	55	314	58	316	58
11 RLA	669	669	183	131	315	61	315	61	319	61	311	65

Table 10.A.2 Percentage of Proficient and Above Across Base Year, 2013, 2014, and 2015

STS	Base		2013		2014		2015
	Overall	Target	Overall	Target	Overall	Target	Overall
2 RLA	40%	40%	36%	37%	29%	30%	24%
3 RLA	37%	37%	37%	36%	19%	20%	22%
4 RLA	35%	35%	36%	35%	28%	29%	24%
5 RLA	28%	29%	30%	29%	25%	28%	26%
6 RLA	29%	28%	33%	32%	32%	32%	27%
7 RLA	30%	31%	36%	38%	30%	30%	30%
8 RLA	33%	35%	33%	35%	31%	31%	26%
9 RLA	32%	33%	32%	33%	31%	31%	24%
10 RLA	36%	37%	36%	37%	29%	29%	27%
11 RLA	30%	30%	30%	30%	36%	38%	32%

Table 10.A.3 Percentage of Advanced Across Base Year, 2013, 2014, and 2015

STS	Base		2013		2014		2015
	Overall	Target	Overall	Target	Overall	Target	Overall
2 RLA	17%	17%	17%	17%	10%	10%	9%
3 RLA	13%	13%	15%	14%	4%	5%	6%
4 RLA	12%	12%	15%	14%	8%	8%	7%
5 RLA	8%	9%	11%	10%	7%	9%	10%
6 RLA	7%	7%	12%	11%	10%	10%	8%
7 RLA	8%	8%	14%	15%	9%	9%	11%
8 RLA	10%	11%	10%	11%	9%	9%	9%
9 RLA	10%	11%	10%	11%	8%	9%	7%
10 RLA	10%	11%	10%	11%	9%	9%	8%
11 RLA	9%	9%	9%	9%	11%	10%	11%

Table 10.A.4 Observed Score Distributions of the STS for RLA Across Base Year, 2013, 2014, and 2015 (Grades Two through Four)

Observed Score Distributions	Gr. 2				Gr. 3				Gr. 4			
	Base	2013	2014	2015	Base	2013	2014	2015	Base	2013	2014	2015
570–600	8	5	4	0	2	1	0	0	0	0	0	0
540–569	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1	0	0	0
510–539	N/A	N/A	N/A	N/A	7	7	0	0	1	N/A	N/A	N/A
480–509	18	26	4	1	23	14	0	0	1	12	0	0
450–479	150	136	10	8	100	89	2	8	19	22	3	3
420–449	385	274	60	25	404	276	20	37	105	171	12	12
390–419	1,320	847	121	103	646	516	45	55	364	328	43	49
360–389	2,187	1,470	234	224	1,382	1,080	153	196	818	635	113	117
330–359	3,203	1,534	382	323	2,032	1,120	255	258	1,007	684	150	157
300–329	2,416	1,649	392	405	1,860	1,276	318	394	955	766	167	187
270–299	1,941	1,484	357	392	1,587	1,097	338	401	846	588	139	184
240–269	1,231	1,090	252	286	812	658	225	240	584	537	117	164
210–239	409	313	98	104	126	113	32	50	181	148	53	55
180–209	29	26	15	8	11	12	0	4	15	12	4	6
150–179	3	0	1	0	0	0	0	0	0	0	2	0

A frequency count of “N/A” indicates that there are no obtainable scale scores within that scale-score range.

Table 10.A.5 Observed Score Distributions of the STS for RLA Across Base Year, 2013, 2014, and 2015 (Grades Five through Seven)

Observed Score Distributions	Gr. 5				Gr. 6				Gr. 7			
	Base	2013	2014	2015	Base	2013	2014	2015	Base	2013	2014	2015
570–600	0	0	0	0	0	1	0	0	0	1	0	0
540–569	0	2	0	0	0	1	0	0	0	1	0	0
510–539	2	3	1	0	1	2	1	0	0	4	0	0
480–509	10	9	1	3	4	12	1	2	3	14	1	2
450–479	27	30	6	9	11	42	9	1	8	34	8	8
420–449	116	83	26	19	50	65	20	13	49	73	19	20
390–419	186	214	39	69	148	125	30	32	122	140	31	31
360–389	359	284	80	66	207	218	70	53	224	169	64	43
330–359	511	463	85	110	369	260	71	56	251	268	76	76
300–329	602	493	106	129	412	336	77	74	327	240	91	79
270–299	583	507	139	126	388	294	84	88	327	238	78	87
240–269	460	452	104	131	271	256	77	77	208	182	71	79
210–239	169	175	50	51	88	97	37	37	66	41	19	25
180–209	23	26	6	17	13	16	5	5	1	9	2	4
150–179	1	2	2	0	0	0	0	2	1	0	0	1

Table 10.A.6 Observed Score Distributions of the STS for RLA Across Base Year, 2014, and 2015 for RLA (Grades Eight through Eleven)

Observed Score Distributions	Gr. 8 Base	Gr. 8 2014	Gr. 8 2015	Gr. 9 Base	Gr. 9 2014	Gr. 9 2015	Gr. 10 Base	Gr. 10 2014	Gr. 10 2015	Gr. 11 Base	Gr. 11 2014	Gr. 11 2015
570–600	0	0	0	0	0	0	0	0	0	0	0	0
540–569	0	0	0	0	0	0	0	0	0	0	0	0
510–539	1	0	0	0	0	0	0	0	0	0	0	0
480–509	8	1	0	2	1	0	1	1	0	0	0	0
450–479	19	9	3	21	8	2	10	1	4	8	0	0
420–449	48	15	14	62	16	8	33	5	12	24	6	7
390–419	93	35	36	166	76	58	110	26	14	34	13	11
360–389	165	58	52	239	121	79	190	63	32	104	36	16
330–359	196	68	74	414	193	183	215	74	52	102	30	20
300–329	203	96	92	399	200	161	262	66	57	121	39	19
270–299	194	68	87	354	140	167	155	53	47	108	16	17
240–269	116	66	74	226	156	110	120	53	39	95	18	20
210–239	34	20	23	79	26	29	62	42	22	49	18	16
180–209	6	2	2	15	17	12	19	3	6	23	7	4
150–179	0	1	0	3	2	1	0	1	0	1	0	1

Appendix 10.B—Historical Comparisons Tables, Test Characteristics

Table 10.B.1 Average Proportion Correct for Operational Test Items Across Base Year, 2013, 2014, and 2015

STS	Base	2013	2014	2015
2 RLA	0.65	0.62	0.59	0.56
3 RLA	0.59	0.59	0.50	0.52
4 RLA	0.57	0.58	0.55	0.52
5 RLA	0.49	0.51	0.49	0.49
6 RLA	0.48	0.52	0.50	0.48
7 RLA	0.53	0.58	0.53	0.53
8 RLA	0.56	0.56	0.52	0.50
9 RLA	0.55	0.55	0.54	0.53
10 RLA	0.57	0.57	0.53	0.54
11 RLA	0.56	0.56	0.55	0.55

Table 10.B.2 Overall IRT *b*-values for Operational Test Items Across Base Year, 2013, 2014, and 2015

STS	Base	2013	2014	2015
2 RLA	-0.59	-0.52	-0.55	-0.52
3 RLA	-0.31	-0.30	-0.26	-0.30
4 RLA	-0.18	-0.18	-0.23	-0.18
5 RLA	0.09	-0.02	0.00	-0.02
6 RLA	0.05	-0.07	-0.02	-0.02
7 RLA	-0.16	-0.23	-0.16	-0.23
8 RLA	-0.26	-0.26	-0.15	-0.15
9 RLA	-0.18	-0.18	-0.19	-0.19
10 RLA	-0.25	-0.25	-0.24	-0.25
11 RLA	-0.22	-0.22	-0.10	-0.22

Table 10.B.3 Average Point-Biserial Correlation for Operational Test Items Across Base Year, 2013, 2014, and 2015

STS	Base	2013	2014	2015
2 RLA	0.43	0.44	0.44	0.41
3 RLA	0.39	0.40	0.36	0.37
4 RLA	0.39	0.42	0.40	0.39
5 RLA	0.36	0.37	0.37	0.38
6 RLA	0.33	0.37	0.38	0.36
7 RLA	0.35	0.39	0.37	0.40
8 RLA	0.38	0.38	0.37	0.36
9 RLA	0.37	0.37	0.37	0.34
10 RLA	0.35	0.35	0.37	0.37
11 RLA	0.35	0.35	0.35	0.37

Table 10.B.4 Score Reliabilities (Cronbach's Alpha) Across Base Year, 2013, 2014, and 2015

STS	Base	2013	2014	2015
2 RLA	0.93	0.94	0.93	0.92
3 RLA	0.91	0.92	0.89	0.90
4 RLA	0.93	0.93	0.93	0.93
5 RLA	0.91	0.91	0.92	0.92
6 RLA	0.89	0.92	0.92	0.91
7 RLA	0.90	0.92	0.92	0.93
8 RLA	0.92	0.92	0.92	0.91
9 RLA	0.91	0.91	0.91	0.90
10 RLA	0.90	0.90	0.91	0.91
11 RLA	0.90	0.90	0.90	0.91

Table 10.B.5 Standard Error of Measurement (SEM) Across Base Year, 2013, 2014, and 2015

STS	Base	2013	2014	2015
	2 RLA	3.34	3.42	3.49
3 RLA	3.51	3.51	3.66	3.63
4 RLA	3.79	3.79	3.86	3.87
5 RLA	3.96	3.94	3.95	3.91
6 RLA	3.95	3.96	3.95	3.97
7 RLA	3.89	3.84	3.88	3.88
8 RLA	3.87	3.87	3.91	3.94
9 RLA	3.84	3.84	3.82	3.85
10 RLA	3.87	3.87	3.87	3.89
11 RLA	3.92	3.92	3.93	3.90