

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



**Standards-based Tests in Spanish  
Technical Report  
Spring 2014 Administration**

**Final Version Submitted February 13, 2015**

**Educational Testing Service**

**Contract No. 5417**



# Table of Contents

Acronyms and Initialisms Used in the <i>STS Technical Report</i> .....	vii
<b>Chapter 1: Introduction</b> .....	<b>1</b>
<b>Background</b> .....	<b>1</b>
<b>Test Purpose</b> .....	<b>1</b>
<b>Test Content</b> .....	<b>1</b>
<b>Intended Population</b> .....	<b>1</b>
<b>Intended Use and Purpose of Test Scores</b> .....	<b>2</b>
<b>Testing Window</b> .....	<b>3</b>
<b>Significant CAASPP Developments in 2014</b> .....	<b>3</b>
Renamed the Program.....	3
Pre-equating of All Results.....	3
Reduced the Number of Tests .....	3
Reduced the Number of Test Versions .....	3
Updated Universal Tools, Designated Supports, and Accommodations (formerly Modifications and Accommodations).....	3
<b>Limitations of the Assessment</b> .....	<b>3</b>
Score Interpretation .....	3
Out-of-Level Testing .....	3
Score Comparison .....	4
<b>Groups and Organizations Involved with the CAASPP System</b> .....	<b>4</b>
State Board of Education.....	4
California Department of Education .....	4
Contractors .....	4
<b>Overview of the Technical Report</b> .....	<b>5</b>
<b>References</b> .....	<b>7</b>
<b>Chapter 2: An Overview of STS Processes</b> .....	<b>8</b>
<b>Item Development</b> .....	<b>8</b>
Item Formats.....	8
Item Specifications.....	8
Item Banking .....	8
Item Refresh Rate.....	9
<b>Test Assembly</b> .....	<b>9</b>
Test Length.....	9
Test Blueprints.....	9
Content Rules and Item Selection.....	9
Psychometric Criteria.....	10
<b>Test Administration</b> .....	<b>10</b>
Test Security and Confidentiality.....	10
Procedures to Maintain Standardization .....	11
<b>Universal Tools, Designated Supports, and Accommodations</b> .....	<b>11</b>
Non-embedded Supports.....	12
Individualized Aids (Previously Called Modifications) .....	12
Special Services Summaries .....	12
<b>Scores</b> .....	<b>12</b>
Aggregation Procedures .....	13
<b>Equating</b> .....	<b>13</b>
Post-Equating .....	13
Pre-Equating.....	14
Calibration.....	14
Scaling.....	14
Scoring Table Production.....	15
<b>References</b> .....	<b>17</b>
<b>Appendix 2.A—STS Items and Estimated Time Chart</b> .....	<b>18</b>
<b>Appendix 2.B—Reporting Clusters</b> .....	<b>19</b>
Reading/Language Arts .....	19
<b>Appendix 2.C—Universal Tools, Designated Supports, and Accommodations for the California Assessment of Student Performance and Progress</b> .....	<b>21</b>
<b>Appendix 2.D—Special Services Summary Tables</b> .....	<b>23</b>
<b>Chapter 3: Item Development</b> .....	<b>58</b>
<b>Rules for Item Development</b> .....	<b>58</b>
Item Specifications.....	58

Expected Item Ratio.....	59
<b>Selection of Item Writers .....</b>	<b>59</b>
Criteria for Selecting Item Writers .....	59
<b>Item Review Process.....</b>	<b>60</b>
Contractor Review .....	60
Content Expert Reviews .....	61
Statewide Pupil Assessment Review Panel .....	64
<b>Field Testing .....</b>	<b>64</b>
Stand-alone Field Testing .....	64
Embedded Field-test Items .....	64
<b>CDE Data Review.....</b>	<b>65</b>
<b>Item Banking.....</b>	<b>65</b>
<b>References .....</b>	<b>67</b>
<b>Chapter 4: Test Assembly .....</b>	<b>68</b>
<b>Test Length.....</b>	<b>68</b>
<b>Rules for Item Selection.....</b>	<b>68</b>
Test Blueprint.....	68
Content Rules and Item Selection.....	68
Replacement Items .....	69
Psychometric Criteria .....	69
Projected Psychometric Properties of the Assembled Tests.....	71
<b>Rules for Item Sequence and Layout .....</b>	<b>71</b>
<b>Reference.....</b>	<b>72</b>
<b>Appendix 4.A—Technical Characteristics.....</b>	<b>73</b>
<b>Appendix 4.B—Cluster Targets.....</b>	<b>76</b>
<b>Chapter 5: Test Administration.....</b>	<b>86</b>
<b>Test Security and Confidentiality.....</b>	<b>86</b>
ETS’s Office of Testing Integrity.....	86
Test Development.....	86
Item and Data Review.....	86
Item Banking.....	87
Transfer of Forms and Items to the CDE .....	87
Security of Electronic Files Using a Firewall .....	87
Printing and Publishing .....	88
Test Administration .....	88
Test Delivery .....	88
Processing and Scoring .....	89
Data Management .....	89
Transfer of Scores via Secure Data Exchange .....	90
Statistical Analysis .....	90
Reporting and Posting Results.....	90
Student Confidentiality .....	90
Student Test Results.....	90
<b>Procedures to Maintain Standardization .....</b>	<b>91</b>
Test Administrators .....	91
Directions for Administration .....	92
CAASPP LEA and Test Site Coordinator Manual .....	92
Test Management System Manuals.....	93
Test Booklets.....	93
<b>Universal Tools, Designated Supports, and Accommodations for Students with Disabilities .....</b>	<b>93</b>
Identification.....	94
Scoring.....	94
<b>Testing Incidents.....</b>	<b>94</b>
Social Media Security Breaches .....	94
<b>Testing Improprieties.....</b>	<b>95</b>
<b>References .....</b>	<b>96</b>
<b>Chapter 6: Performance Standards .....</b>	<b>97</b>
<b>Background .....</b>	<b>97</b>
<b>Standard-Setting Procedure.....</b>	<b>97</b>
Development of Competencies Lists.....	98
<b>Standard-Setting Methodology .....</b>	<b>99</b>
<b>Results .....</b>	<b>100</b>
<b>References .....</b>	<b>101</b>

<b>Chapter 7: Scoring and Reporting</b> .....	<b>102</b>
<b>Procedures for Maintaining and Retrieving Individual Scores</b> .....	<b>102</b>
Scoring and Reporting Specifications .....	102
Scanning and Scoring.....	103
<b>Types of Scores and Subscores</b> .....	<b>103</b>
Raw Score .....	103
Subscore.....	103
Scale Score.....	103
Performance Levels .....	103
<b>Score Verification Procedures</b> .....	<b>104</b>
Scoring Key Verification Process .....	104
<b>Overview of Score Aggregation Procedures</b> .....	<b>104</b>
Individual Scores.....	104
Group Scores.....	107
<b>Reports Produced and Scores for Each Report</b> .....	<b>108</b>
Types of Score Reports .....	108
Score Report Contents .....	109
Score Report Applications.....	109
<b>Criteria for Interpreting Test Scores</b> .....	<b>110</b>
<b>Criteria for Interpreting Score Reports</b> .....	<b>110</b>
<b>References</b> .....	<b>111</b>
<b>Appendix 7.A—Scale Score Distribution Tables</b> .....	<b>112</b>
<b>Appendix 7.B—Demographic Summaries</b> .....	<b>113</b>
<b>Appendix 7.C—Types of Score Reports</b> .....	<b>134</b>
<b>Chapter 8: Analyses</b> .....	<b>136</b>
<b>Samples Used for the Analyses</b> .....	<b>136</b>
<b>Classical Item Analyses</b> .....	<b>137</b>
Multiple-Choice Items .....	137
<b>Reliability Analyses</b> .....	<b>138</b>
Subgroup Reliabilities and SEMs.....	139
Conditional Standard Errors of Measurement.....	139
<b>Decision Classification Analyses</b> .....	<b>140</b>
<b>Validity Evidence</b> .....	<b>141</b>
Purpose of the STS.....	142
The Constructs to Be Measured .....	142
Interpretations and Uses of the Scores Generated .....	143
Intended Test Population(s).....	143
Validity Evidence Collected.....	143
Evidence Based on Response Processes .....	145
Evidence Based on Internal Structure.....	146
Evidence Based on Consequences of Testing.....	147
<b>IRT Analyses</b> .....	<b>147</b>
Post-equating.....	147
Pre-Equating.....	148
Summaries of Scaled IRT <i>b</i> -values.....	148
Evaluation of Pre-equating.....	149
Equating Results.....	149
<b>Differential Item Functioning Analyses</b> .....	<b>149</b>
<b>References</b> .....	<b>151</b>
<b>Appendix 8.A—Classical Analyses</b> .....	<b>153</b>
<b>Appendix 8.B—Reliability Analyses</b> .....	<b>159</b>
<b>Appendix 8.C—IRT Analyses</b> .....	<b>174</b>
Scaling and Equating Results .....	175
<b>Chapter 9: Quality Control Procedures</b> .....	<b>185</b>
<b>Quality Control of Item Development</b> .....	<b>185</b>
Item Specifications.....	185
Item Writers.....	185
Internal Contractor Reviews.....	185
Assessment Review Panel Review .....	186
Statewide Pupil Assessment Review Panel Review .....	186
Data Review of Field-tested Items .....	186
<b>Quality Control of the Item Bank</b> .....	<b>187</b>
<b>Quality Control of Test Form Development</b> .....	<b>187</b>

<b>Quality Control of Test Materials .....</b>	<b>188</b>
Collecting Test Materials.....	188
Processing Test Materials.....	188
<b>Quality Control of Scanning.....</b>	<b>188</b>
<b>Quality Control of Image Editing.....</b>	<b>189</b>
<b>Quality Control of Answer Document Processing and Scoring.....</b>	<b>189</b>
Accountability of Answer Documents.....	189
Processing of Answer Documents.....	189
Scoring and Reporting Specifications.....	190
Storing Answer Documents.....	190
<b>Quality Control of Psychometric Processes.....</b>	<b>190</b>
Score Key Verification Procedures.....	190
Quality Control of Item Analyses and the Equating Process.....	190
Score Verification Process.....	192
Year-to-Year Comparison Analyses.....	192
Offloads to Test Development.....	192
<b>Quality Control of Reporting.....</b>	<b>192</b>
Electronic Reporting.....	193
Excluding Student Scores from Summary Reports.....	193
<b>Reference.....</b>	<b>194</b>
<b>Chapter 10: Historical Comparisons.....</b>	<b>195</b>
<b>Base Year Comparisons.....</b>	<b>195</b>
<b>Examinee Performance.....</b>	<b>196</b>
<b>Test Characteristics.....</b>	<b>196</b>
<b>Appendix 10.A—Historical Comparisons Tables, Examinee Performance.....</b>	<b>197</b>
<b>Appendix 10.B—Historical Comparisons Tables, Test Characteristics.....</b>	<b>200</b>

## 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.....	21
Table 2.D.1 Special Services Summary for RLA, Grades Two and Three.....	23
Table 2.D.2 Special Services Summary for RLA, Grades Four and Five.....	30
Table 2.D.3 Special Services Summary for RLA, Grades Six and Seven.....	37
Table 2.D.4 Special Services Summary for RLA, Grades Eight and Nine.....	44
Table 2.D.5 Special Services Summary for RLA, Grades Ten and Eleven.....	51
Table 4.1 Statistical Targets for STS Test Assembly.....	70
Table 4.A.1 Summary of 2014 STS Projected Technical Characteristics.....	73
Table 4.A.2 Summary of 2014 STS Projected Statistical Attributes.....	73
Table 7.1 Mean and Standard Deviation of Raw and Scale Scores for the STS (Overall Population).....	105
Table 7.2 Mean and Standard Deviation of Raw and Scale Scores for the STS (Target Population).....	105
Table 7.3 Mean and Standard Deviation of Raw and Scale Scores for the STS (Optional Population).....	106
Table 7.4 Percentages of Examinees in Each Performance Level.....	106
Table 7.5 Subgroup Definitions.....	107
Table 7.6 Types of STS Reports.....	108
Table 7.A.1 Distribution of STS Scale Scores for RLA, Grades Two through Eleven (Overall Population).....	112
Table 7.A.2 Distribution of STS Scale Scores for RLA, Grades Two through Eleven (Target Population).....	112
Table 7.B.1 Demographic Summary for RLA, Grade Two (Overall Population).....	113
Table 7.B.2 Demographic Summary for RLA, Grade Two (Target Population).....	115
Table 7.B.3 Demographic Summary for RLA, Grade Three (Overall Population).....	116
Table 7.B.4 Demographic Summary for RLA, Grade Three (Target Population).....	117
Table 7.B.5 Demographic Summary for RLA, Grade Four (Overall Population).....	118
Table 7.B.6 Demographic Summary for RLA, Grade Four (Target Population).....	119
Table 7.B.7 Demographic Summary for RLA, Grade Five (Overall Population).....	120
Table 7.B.8 Demographic Summary for RLA, Grade Five (Target Population).....	121
Table 7.B.9 Demographic Summary for RLA, Grade Six (Overall Population).....	122
Table 7.B.10 Demographic Summary for RLA, Grade Six (Target Population).....	123
Table 7.B.11 Demographic Summary for RLA, Grade Seven (Overall Population).....	124
Table 7.B.12 Demographic Summary for RLA, Grade Seven (Target Population).....	125
Table 7.B.13 Demographic Summary for RLA, Grade Eight (Overall Population).....	126
Table 7.B.14 Demographic Summary for RLA, Grade Eight (Target Population).....	127
Table 7.B.15 Demographic Summary for RLA, Grade Nine (Overall Population).....	128
Table 7.B.16 Demographic Summary for RLA, Grade Nine (Target Population).....	129
Table 7.B.17 Demographic Summary for RLA, Grade Ten (Overall Population).....	130
Table 7.B.18 Demographic Summary for RLA, Grade Ten (Target Population).....	131

Table 7.B.19 Demographic Summary for RLA, Grade Eleven (Overall Population).....	132
Table 7.B.20 Demographic Summary for RLA, Grade Eleven (Target Population).....	133
Table 7.C.1 Score Reports Reflecting STS Results.....	134
Table 8.1 Mean and Median Proportion Correct and Point-Biserial Correlation.....	137
Table 8.2 Reliabilities and SEMs for the STS .....	139
Table 8.3 Scale Score CSEM at Performance-level Cut Points .....	140
Table 8.4 Original Year of Administration for the STS .....	145
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.....	153
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.....	154
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.....	156
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.....	157
Table 8.B.1 Subscore Reliabilities and Intercorrelations for RLA.....	159
Table 8.B.2 Reliabilities and SEMs by Gender .....	160
Table 8.B.3 Reliabilities and SEMs by Economic Status .....	160
Table 8.B.4 Reliabilities and SEMs by Special Services .....	161
Table 8.B.5 Reliabilities and SEMs by Attendance in U.S. Schools.....	161
Table 8.B.6 Reliabilities and SEMs by EL Program Participation.....	162
Table 8.B.7 Subscore Reliabilities and SEM for RLA by Gender/Economic Status .....	163
Table 8.B.8 Subscore Reliabilities and SEM for RLA by Special Services/Attendance in U.S. Schools .....	165
Table 8.B.9 Subscore Reliabilities and SEM for RLA by EL Program Participation .....	167
Table 8.B.10 Reliability of Classification for RLA, Grade Two.....	170
Table 8.B.11 Reliability of Classification for RLA, Grade Three .....	170
Table 8.B.12 Reliability of Classification for RLA, Grade Four.....	170
Table 8.B.13 Reliability of Classification for RLA, Grade Five.....	171
Table 8.B.14 Reliability of Classification for RLA, Grade Six .....	171
Table 8.B.15 Reliability of Classification for RLA, Grade Seven .....	171
Table 8.B.16 Reliability of Classification for RLA, Grade Eight .....	172
Table 8.B.17 Reliability of Classification for RLA, Grade Nine .....	172
Table 8.B.18 Reliability of Classification for RLA, Grade Ten .....	172
Table 8.B.19 Reliability of Classification for RLA, Grade Eleven .....	173
Table 8.C.1 IRT $b$ -values for RLA, Grade Six .....	174
Table 8.C.2 IRT $b$ -values for RLA, Grade Eight.....	174
Table 8.C.3 Distribution of IRT Difficulty ( $b$ -values) for RLA (operational items).....	174
Table 8.C.4 New Conversion for RLA, Grade Two .....	175
Table 8.C.5 New Conversion for RLA, Grade Three.....	176
Table 8.C.6 New Conversion for RLA, Grade Four.....	177
Table 8.C.7 New Conversion for RLA, Grade Five .....	178
Table 8.C.8 New Conversion for RLA, Grade Six .....	179
Table 8.C.9 New Conversion for RLA, Grade Seven .....	180
Table 8.C.10 New Conversion for RLA, Grade Eight.....	181
Table 8.C.11 New Conversion for RLA, Grade Nine.....	182
Table 8.C.12 New Conversion for RLA, Grade Ten .....	183
Table 8.C.13 New Conversion for RLA, Grade Eleven .....	184
Table 10.1 Base Years for STS .....	195
Table 10.A.1 Number of Examinees Tested, Scale Score Means, and Standard Deviations of STS Across Base Year, 2012, 2013, and 2014 .....	197
Table 10.A.2 Percentage of Proficient and Above and Percentage of Advanced Across Base Year, 2012, 2013, and 2014 .....	197
Table 10.A.3 Observed Score Distributions of the STS Across Base Year, 2012, 2013, and 2014 (Grades Two through Four).....	198
Table 10.A.4 Observed Score Distributions of the STS Across Base Year, 2012, 2013, and 2014 (Grades Five through Seven).....	198
Table 10.A.5 Observed Score Distributions of the STS Across Base Year and 2014 for RLA (Grades Eight through Eleven).....	199
Table 10.B.1 Average Proportion Correct for Operational Test Items Across Base Year, 2012, 2013, and 2013.....	200
Table 10.B.2 Overall IRT $b$ -values for Operational Test Items Across Base Year, 2012, 2013, and 2014 .....	200
Table 10.B.3 Average Point-Biserial Correlation for Operational Test Items Across Base Year, 2012, 2013, and 2014 ..	201
Table 10.B.4 Score Reliabilities (Cronbach's Alpha) and Standard Error of Measurement (SEM) Across Base Year, 2012, 2013, and 2014 .....	201

## Figures

Figure 3.1 The ETS Item Development Process for the CAASPP System .....	58
Figure 4.A.1 Plots of Target Information Function and Projected Information for Total Test and Linking Set for RLA.....	74
Figure 4.B.1 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Two .....	76
Figure 4.B.2 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Three .....	77
Figure 4.B.3 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Four .....	78
Figure 4.B.4 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Five .....	79
Figure 4.B.5 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Six.....	80
Figure 4.B.6 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Seven.....	81
Figure 4.B.7 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Eight.....	82
Figure 4.B.8 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Nine .....	83
Figure 4.B.9 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Ten .....	84
Figure 4.B.10 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Eleven.....	85
Figure 6.1 Bookmark Standard-setting Process for the STS .....	99
Figure 8.1 Decision Accuracy for Achieving a Performance Level.....	141
Figure 8.2 Decision Consistency for Achieving a Performance Level .....	141

**Acronyms and Initialisms Used in the STS Technical Report**

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



# 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 (ELA), 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 its 2014 administration, the CAASPP System had four components for the paper-pencil tests:

- CSTs, produced for California public schools to assess the California content standards for science in grades five, eight, and ten
- CMA, 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, produced for students with an IEP and who have significant cognitive disabilities in grades two through eleven and are not able to take the CSTs with accommodations and/or non-embedded accessibility supports or the CMA with accommodations
- STS, an assessment of students’ achievement of California’s content standards for Spanish-speaking English learners that is administered as the CAASPP System’s designated primary language test (DPLT)

## 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 (RLA) 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 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 are required to also take the CSTs and/or CMA appropriate to their grade level.

The STS 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. 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. (Note: These are *not* the students who took the STS for Non-ELs in Dual-immersion Programs. In 2014, the STS could be purchased by 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.)

The number of examinees taking the grade-level STS varies significantly across different grade levels, from approximately 200 in grade eleven (RLA) 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 three primary purposes, described as follows (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. (a) (1) First and foremost, provide information on the academic status and progress of individual pupils to those pupils, their parents, and their teachers. This information should be designed to assist in the improvement of teaching and learning in California public classrooms. The Legislature recognizes that, in addition to statewide assessments that will occur as specified in this chapter, school districts will conduct additional ongoing pupil diagnostic assessment and provide information regarding pupil performance based on those assessments on a regular basis to parents or guardians and schools. The Legislature further recognizes that local diagnostic assessment is a primary mechanism through which academic strengths and weaknesses are identified.”

“60602. (a) (4) Provide information to pupils, parents or guardians, teachers, schools, and school districts 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. (c) It is the intent of the Legislature that parents, classroom teachers, other educators, governing board members of school districts, and the public be involved, in an active and ongoing basis, in the design and implementation of the statewide pupil assessment program and the development of assessment instruments.”

“60602. (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 Standardized Testing and Reporting Program become open and transparent to teachers, parents, and pupils, to assist all the 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 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]*).

## Significant CAASPP Developments in 2014

### Renamed the Program

The paper-pencil tests administered in 2014 are a component of the CAASPP System.

### Pre-equating of All Results

Because intact test forms were reused, raw-score-to-scale-score conversion tables were developed before tests were administered and used on these tests. This process was used on all STS for RLA forms.

### Reduced the Number of Tests

Because California is in transition to the new assessment system, the number of non-computer administered tests is reduced to only include grade-level RLA.

### Reduced the Number of Test Versions

The number of STS for RLA versions available for administration was reduced to two (grades eight, ten, and eleven) or three (grades two through seven and nine).

### Updated Universal Tools, Designated Supports, and Accommodations (formerly Modifications and Accommodations)

Students were permitted the use of universal tools, designated supports, and accommodations as outlined in 5 CCR Section 853.5(b), (d), (f), and (h).

## 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 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 or any test in the CAASPP System; all students are required to take the 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, 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 grade three RLA tests; it would not be appropriate to compare scores obtained on the grade three RLA test with those obtained on the grade four RLA test. 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 responsible for assuring the compliance with programs that meet the requirement of the federal Elementary and Secondary Education Act (ESEA) and the state's Public School Accountability Act (PSAA) and for reporting CAPA results in terms of the Adequate Yearly Progress, and, in previous years, Academic Performance Index for the CSTs, CMA, and CAPA; these measure the academic performance and growth of schools on a variety of academic metrics.

### **California Department of Education**

The CDE 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 CDE adopts textbooks for kindergarten through grade eight, adopts regulations to implement legislation, and has the authority to grant waivers of the *EC*.

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.

### **Contractors**

#### **Educational Testing Service**

The CDE and the SBE contract with ETS to develop, administer, and report the CAASPP assessments. As the prime contractor, ETS has overall responsibility for working with the CDE to implement and maintain an effective assessment system and to coordinate the work of ETS and its subcontractor Pearson. 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 Management System for orders of materials and pre-identification services; and
- Completion of all psychometric activities.

### **Pearson**

ETS also monitors and manages the work of Pearson, subcontractor to ETS for the CAASPP System. Activities conducted by Pearson include the following:

- 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, including performance scoring of the writing 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 administered in spring 2014. 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 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 items before the 2014 administration—in 2014, 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 forms reused in 2014.
- Chapter 5 presents the processes involved in the actual administration of the 2014 STS 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 and includes a discussion of quick-turnaround reporting.
- Chapter 8 summarizes the results of the test- and item-level analyses performed during the spring 2014 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 performance-level classifications, and the procedures designed to ensure the validity of STS score uses and interpretations. Also discussed in

this chapter are item response theory (IRT), STS 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.
- Chapter 10 presents historical comparisons of various item- and test-level results for the past three years (grades two through seven) or one year (grades eight through eleven) 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](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>

## Chapter 2: An Overview of STS Processes

---

This chapter provides an overview of the processes involved in a typical test development and administration cycle for the STS; these processes are similar to those undertaken to develop the CSTs. Also described are the specifications maintained by ETS to implement each of those processes. In 2014, forms from previous test administrations from different years were reused. 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 contain four-option multiple-choice items.

#### Item Specifications

There was no new item development in 2014. Prior to the 2014 administration, the STS items were developed to measure California content standards 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 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 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 measure the content standards in the same way. To achieve this, the item specifications provided detailed information to item writers who are developing items for the STS.

The items selected for each STS 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 58.

#### 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 61; and the Statewide Pupil Assessment Review (SPAR) panel, described in Chapter 3, starting on page 64.

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 65 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 65 in Chapter 3.

### **Item Refresh Rate**

Prior to form reuse using intact forms in the 2014 administration, the item utilization plan required that each year, a certain percentage of items on an operational form were refreshed; these items remained in the item bank for future use.

## **Test Assembly**

### **Test Length**

The number of operational items in each STS 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 68 in Chapter 4.

Each STS also includes six field-test items in addition to the operational items. Although there was no new item development for the 2014 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 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 items to conform to the SBE-approved California content standards and test blueprints. The test blueprints for the STS can be found on the CDE STAR STS Blueprints Web page at <http://www.cde.ca.gov/ta/tg/sr/stsblueprints.asp>.

Although the test blueprints specify the number of items at the individual standard level, scores for the STS items are grouped into subcontent areas referred to as "reporting clusters." For each STS reporting cluster, the percentage of questions correctly answered is reported on a student's score report. A description of the STS 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 administration. (See Table 8.4 on page 145 for administration years.) Prior to the 2014 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 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 68.

### **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 forms and the projected characteristics of the assembled forms are presented starting from page 69 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 71 in Chapter 4.

All the forms in the 2014 STS administration were used in prior operational test administrations. See Table 8.4 on page 145 for the list containing the administration in which each STS was originally administered. Finally, see page 69 in Chapter 4 for a description of the item replacement process.

### **Test Administration**

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

### **Test Security and Confidentiality**

All tests within the CAASPP System are secure documents. For the STS 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 86.

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 86.

- 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 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, as described in this section.

### Test Administrators

The STS 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 LEAs who are central to the processes include LEA CAASPP Coordinators, CAASPP Test Site Coordinators, test examiners, proctors, and scribes. The responsibilities of each of the staff members are included in the *CAASPP LEA and Test Site Coordinator Manual* (CDE, 2014); see page 92 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 examiners to administer the STS to students to be followed exactly so that all students have an equal opportunity to demonstrate their academic achievement (See page 92 in Chapter 5 for more information.)

***CAASPP LEA and Test Site Coordinator Manual***—Test administration procedures for LEA CAASPP Coordinators and CAASPP Test Site Coordinators (See page 92 in Chapter 5 for more information.)

***Test Management System manuals***—Instructions for the Web-based modules that allow LEA CAASPP Coordinators to set up test administrations, order materials, and submit and correct student Pre-ID data; every module has its own user manual with detailed instructions on how to use the Test Management System (See page 93 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 ELs. Most students with IEPs take the STS under standard conditions. However, some students with IEPs may need assistance when taking the STS. 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.” 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 in 2014 and were mapped to STS 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 for the STS is to enable the students to take the assessments, rather than give them an advantage over other students or to artificially inflate their scores.

### **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 (Previously Called Modifications)**

Individualized aids, previously called modifications, 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.

### **Special Services Summaries**

The percentage of students using various universal tools, designated supports, and accommodations during the 2014 administration of the STS is presented in Appendix 2.D, which starts on page 23. The data are organized into five 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 target and nontarget STS students are presented in the second section, and the results for students in special education and not in special education are presented in the third section. The fourth 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 final section presents the results for various categories based on EL program participation. The information within each section is presented for the relevant grades.

## **Scores**

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

Total test raw scores on each STS are converted to three-digit scale scores using the pre-equating process described starting on page 14. STS 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 scores are found in Chapter 7, which starts on page 102.

## Aggregation Procedures

In order to provide meaningful results to the stakeholders, STS 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 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 as a designated EL
5. Did not test out of level (grade inappropriate)

### Individual Scores

Table 7.1 through Table 7.3, starting on page 105 in Chapter 7, offer summary statistics for individual scores aggregated at the state level, describing overall student performance on each STS for the total, target, and nontarget STS student populations respectively. Included in the tables 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 7.4 on page 106 presents the percentages of STS target students in each performance level.

### Demographic Subgroup Scores

Statistics summarizing STS student performance by content area and grade for selected groups of students are provided in Table 7.B.1 through Table 7.B.20, starting on page 113 in Appendix 7.B, for overall and target STS students, respectively. 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 7.5 on page 107 provides definitions for the demographic groups included in the tables.

## Equating

### Post-Equating

Prior to the 2013 administration, the STS 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 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 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.

## Pre-Equating

During the 2014 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.

The original conversion tables from previous administrations of the STS for RLA in grades two through eleven except for grades six and eight are directly applied to the current administration. For the STS for RLA in grades six and eight, in which replacement items are used—one replacement item in grade six and five replacement items in grade eight—conversion tables are newly generated using the scoring table production described below. The item parameters used for true-score equating are post-equating item parameters (from the reused forms) for the unchanged items and the post-equated parameters from the most recent administration for the replacement or edited items.

Table 8.4 on page 145 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

Prior to the 2013 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 (ICCs) 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,

$\theta_j$  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  $\theta_j$ , and

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

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, 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  $\theta$ , and

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

For each integer score  $\xi_n$  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:

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

where,

$\theta$  represents student ability.

The slope and intercept for each STS 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.

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

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

where,

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

$\theta_{\text{basic}}$  represents theta cut score for basic on the base scale.

Complete raw-score-to-scale-score conversion tables for the STS are presented in Table 8.C.4 through Table 8.C.13 in Appendix 8.C, starting on page 175. 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, 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**

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

\* Numbers indicate grade-level tests.

## References

- California Department of Education. (2014). *2014 LEA CAASPP and test site coordinator manual*. Sacramento, CA. Retrieved from [http://caaspp.org/rsc/pdfs/CAASPP.coord\\_man.2014.pdf](http://caaspp.org/rsc/pdfs/CAASPP.coord_man.2014.pdf)
- California Department of Education. (2014). *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 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
<b>Grade 2</b>	Total No. of Items	71			
	Time	<b>150</b>	50	50	50
<b>Grade 3</b>	Total No. of Items	71			
	Time	<b>150</b>	50	50	50
<b>Grade 4</b>	Total No. of Items	81			–
	Time	<b>170</b>	85	85	–
<b>Grade 5</b>	Total No. of Items	81			–
	Time	<b>170</b>	85	85	–
<b>Grade 6</b>	Total No. of Items	81			–
	Time	<b>170</b>	85	85	–
<b>Grade 7</b>	Total No. of Items	81			–
	Time	<b>170</b>	85	85	–
<b>Grade 8</b>	Total No. of Items	81			–
	Time	<b>170</b>	85	85	–
<b>Grade 9</b>	Total No. of Items	81			–
	Time	<b>170</b>	85	85	–
<b>Grade 10</b>	Total No. of Items	81			–
	Time	<b>170</b>	85	85	–
<b>Grade 11</b>	Total No. of Items	81			–
	Time	<b>170</b>	85	85	–

## Appendix 2.B—Reporting Clusters

### 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)***Reading*

Word Analysis and Vocabulary Development	11 items
Reading Comprehension	18 items
Literary Response and Analysis	13 items

*Writing*

Written Conventions	16 items
Writing Strategies	17 items

**Reading/Language Arts Standards Test (Grade Eight)***Reading*

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

*Writing*

Written Conventions	16 items
Writing Strategies	17 items

**Reading/Language Arts Standards Test (Grade Nine)***Reading*

Word Analysis and Vocabulary Development	8 items
Reading Comprehension	18 items
Literary Response and Analysis	16 items

*Writing*

Written Conventions	13 items
Writing Strategies	20 items

**Reading/Language Arts Standards Test (Grade Ten)***Reading*

Word Analysis and Vocabulary Development	8 items
Reading Comprehension	18 items
Literary Response and Analysis	16 items

*Writing*

Written Conventions	13 items
Writing Strategies	20 items

**Reading/Language Arts Standards Test (Grade Eleven)***Reading*

Word Analysis and Vocabulary Development	8 items
Reading Comprehension	19 items
Literary Response and Analysis	17 items

*Writing*

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**

Option	(U) Universal Tool   (D) Designated Support   (A) Accommodation	
<b>Answer Document Section A3—Accommodations and Modifications</b>		
<b>B</b>	Pupil marks in paper-pencil test booklet (other than responses including highlighting)	U
<b>C</b>	Scribe (previously known as “Essay responses dictated orally, in Manually Coded English, or in American Sign Language to a scribe, audio recorder, or speech-to-text converter” or “Student marks responses in test booklet and responses are transferred to a scorable answer document by an employee of the school, district, or nonpublic school” or “Student dictates multiple-choice question responses orally, or in Manually Coded English to a scribe, audio recorder, or speech-to-text converter for selected-response items”)	A
<b>F</b>	Alternate Response Options includes adapted keyboards, large keyboards, StickyKeys, MouseKeys, FilterKeys, adapted mouse, touch screen, head wand, and switches. (previously known as “Assistive device that does not interfere with the independent work of the student on the multiple choice and/or essay responses [writing portion of the test]”)	–
<b>G</b>	Braille (paper-pencil tests)	A
<b>H</b>	Large-print versions of a paper-pencil test (as available)	A
<b>J, K</b>	Breaks (previously known as “Extended Time” or “Test over more than one day for a test or test part to be administered in a single sitting” or “supervised breaks within a section of the test”)	U
<b>L</b>	Administration of the test to the pupil at the most beneficial time of day	A
<b>M</b>	Separate Setting (previously known as “Test individual student separately, provided that a test examiner directly supervises the student” or “Test student in a small group setting” or “Test administered at home or in hospital by test examiner”)	A
<b>N</b>	Bilingual Dictionary	–
<b>N</b>	English Dictionary	–
<b>N</b>	Thesaurus	–
<b>O</b>	American Sign Language	A
<b>V</b>	Assistive device that interferes with the independent work of the student on the multiple-choice and/or essay responses	Unmapped
<b>W</b>	Unmapped	Unmapped
<b>X</b>	Abacus	A
<b>Y</b>	Leave blank	Unmapped
<b>Z</b>	Read Aloud (previously known as “Test questions and answer options read aloud to pupil or used audio CD presentation – excluding passages”)	A
<b>Mark Nothing</b>	Color Overlay (previously known as “Colored overlay, mask, or other means to maintain visual attention”)	U
<b>Mark Nothing</b>	Magnification (previously known as “Visual magnifying equipment”)	D

Option	(U) Universal Tool   (D) Designated Support   (A) Accommodation	
<b>Mark Nothing</b>	Noise buffers (e.g., individual carrel or study enclosure or noise-cancelling headphones)	D
<b>Mark Nothing</b>	Scratch Paper	U
<b>Mark Nothing</b>	Simplified or clarified test administration directions (does not apply to test questions)	U
<b>Mark Nothing</b>	Special lighting or acoustics, assistive devices (specific devices may require CAASPP contractor certification), and/or special or adaptive furniture	D
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.

**Note:** The use of additional accessibility supports can be requested.

## 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.

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

Special Services Summary for RLA, Grades Two and Three				
All Students Tested	Grade 2	Pct. of Total	Grade 3	Pct. of Total
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	1	0.04%	0	0.00%
J: Breaks (Tested over more than one day)	0	0.00%	6	0.32%
K: Breaks (Had supervised breaks)	10	0.36%	8	0.42%
L: Administered at most beneficial time of day	1	0.04%	1	0.05%
M: Separate setting	6	0.21%	1	0.05%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	1	0.04%	0	0.00%
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	1	0.05%
Z: Read aloud	8	0.28%	4	0.21%
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	16	0.57%	10	0.53%
<i>Any universal tool, desig. support, or accommodation</i>	20	0.71%	12	0.63%
<i>No Universal tool, desig. support, or accommodation</i>	2,788	99.29%	1,888	99.37%
Target Students Tested	Grade 2	Pct. of Total	Grade 3	Pct. of Total
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	1	0.08%	0	0.00%
J: Breaks (Tested over more than one day)	0	0.00%	4	0.48%
K: Breaks (Had supervised breaks)	1	0.08%	4	0.48%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%

<b>Special Services Summary for RLA, Grades Two and Three</b>				
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	7	0.59%	3	0.36%
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	6	0.51%	5	0.60%
<i>Any</i> Universal tool, desig. support, or accommodation	9	0.76%	7	0.84%
<i>No</i> Universal tool, desig. support, or accommodation	1,173	99.24%	831	99.16%
<b>Nontarget (Optional) Students Tested</b>	<b>Grade 2</b>	<b>Pct. of Total</b>	<b>Grade 3</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	2	0.19%
K: Breaks (Had supervised breaks)	9	0.55%	4	0.38%
L: Administered at most beneficial time of day	1	0.06%	1	0.09%
M: Separate setting	6	0.37%	1	0.09%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	1	0.06%	0	0.00%
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	1	0.09%
Z: Read aloud	1	0.06%	1	0.09%
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	10	0.62%	5	0.47%
<i>Any</i> Universal tool, desig. support, or accommodation	11	0.68%	5	0.47%
<i>No</i> Universal tool, desig. support, or accommodation	1,615	99.32%	1,057	99.53%
<b>Students Not in Special Education</b>	<b>Grade 2</b>	<b>Pct. of Total</b>	<b>Grade 3</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	1	0.04%	0	0.00%
J: Breaks (Tested over more than one day)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%

<b>Special Services Summary for RLA, Grades Two and Three</b>				
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	1	0.04%	0	0.00%
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	0	0.00%
Z: Read aloud	1	0.04%	2	0.11%
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	0.04%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	3	0.11%	2	0.11%
<i>No</i> Universal tool, desig. support, or accommodation	2,688	99.89%	1,831	99.89%
<b>Students in Special Education</b>	<b>Grade 2</b>	<b>Pct. of Total</b>	<b>Grade 3</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	6	8.96%
K: Breaks (Had supervised breaks)	10	8.55%	8	11.94%
L: Administered at most beneficial time of day	1	0.85%	1	1.49%
M: Separate setting	6	5.13%	1	1.49%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	0	0.00%	0	0.00%
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	1	1.49%
Z: Read aloud	7	5.98%	2	2.99%
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	15	12.82%	10	14.93%
<i>Any</i> Universal tool, desig. support, or accommodation	17	14.53%	10	14.93%
<i>No</i> Universal tool, desig. support, or accommodation	100	85.47%	57	85.07%
<b>Students in U.S. Schools &lt; 12 Months</b>	<b>Grade 2</b>	<b>Pct. of Total</b>	<b>Grade 3</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	1	0.18%	0	0.00%
J: Breaks (Tested over more than one day)	0	0.00%	4	0.83%
K: Breaks (Had supervised breaks)	0	0.00%	4	0.83%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	0	0.00%	0	0.00%

<b>Special Services Summary for RLA, Grades Two and Three</b>				
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	0	0.00%
Z: Read aloud	7	1.29%	2	0.42%
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	6	1.10%	4	0.83%
<i>Any</i> Universal tool, desig. support, or accommodation	8	1.47%	6	1.25%
<i>No</i> Universal tool, desig. support, or accommodation	536	98.53%	475	98.75%
<b>Students in U.S. Schools ≥ 12 Months</b>	<b>Grade 2</b>	<b>Pct. of Total</b>	<b>Grade 3</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	2	0.14%
K: Breaks (Had supervised breaks)	10	0.44%	4	0.28%
L: Administered at most beneficial time of day	1	0.04%	1	0.07%
M: Separate setting	6	0.27%	1	0.07%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	1	0.04%	0	0.00%
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	1	0.07%
Z: Read aloud	1	0.04%	2	0.14%
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	10	0.44%	6	0.42%
<i>Any</i> Universal tool, desig. support, or accommodation	12	0.53%	6	0.42%
<i>No</i> Universal tool, desig. support, or accommodation	2,252	99.47%	1,413	99.58%
<b>EL Program: Primary Language Instruction and ELD and/or SDAIE Instruction</b>	<b>Grade 2</b>	<b>Pct. of Total</b>	<b>Grade 3</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	0	0.00%	0	0.00%
G: Braille	0	0.00%	0	0.00%
H: Large-print versions of a paper-pencil test	1	0.06%	0	0.00%
J: Breaks (Tested over more than one day)	0	0.00%	6	0.53%
K: Breaks (Had supervised breaks)	3	0.18%	7	0.62%
L: Administered at most beneficial time of day	1	0.06%	1	0.09%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	1	0.06%	0	0.00%

<b>Special Services Summary for RLA, Grades Two and Three</b>				
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	1	0.09%
Z: Read aloud	7	0.42%	1	0.09%
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	10	0.60%	8	0.71%
<i>Any</i> Universal tool, desig. support, or accommodation	12	0.72%	8	0.71%
<i>No</i> Universal tool, desig. support, or accommodation	1,664	99.28%	1,120	99.29%
<b>EL Program: ELD Instruction Only</b>	<b>Grade 2</b>	<b>Pct. of Total</b>	<b>Grade 3</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	21	100.00%	29	100.00%
<b>EL Program: SDAIE Instruction Only</b>	<b>Grade 2</b>	<b>Pct. of Total</b>	<b>Grade 3</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	0	0.00%	0	0.00%

<b>Special Services Summary for RLA, Grades Two and Three</b>				
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	0	0.00%
Z: Read aloud	1	0.46%	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	1	0.46%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	215	99.54%	171	100.00%
<b>EL Program: ELD Instruction and SDAIE Instruction but Not Primary Language Instruction</b>	<b>Grade 2</b>	<b>Pct. of Total</b>	<b>Grade 3</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	1	0.13%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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%	3	0.62%
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%	1	0.21%
<i>Any</i> Universal tool, desig. support, or accommodation	1	0.13%	3	0.62%
<i>No</i> Universal tool, desig. support, or accommodation	758	99.87%	481	99.38%
<b>EL Program: Other EL Instructional Services</b>	<b>Grade 2</b>	<b>Pct. of Total</b>	<b>Grade 3</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	0	0.00%	0	0.00%

<b>Special Services Summary for RLA, Grades Two and Three</b>				
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	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	47	100.00%	28	100.00%
<b>EL Program: None (EL only)</b>	<b>Grade 2</b>	<b>Pct. of Total</b>	<b>Grade 3</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	6	9.68%	1	2.38%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	6	9.68%	1	2.38%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	6	9.68%	1	2.38%
<i>Any</i> Universal tool, desig. support, or accommodation	6	9.68%	1	2.38%
<i>No</i> Universal tool, desig. support, or accommodation	56	90.32%	41	97.62%

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

<b>Special Services Summary for RLA, Grades Four and Five</b>				
<b>All Students Tested</b>	<b>Grade 4</b>	<b>Pct. of Total</b>	<b>Grade 5</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	3	0.26%	0	0.00%
K: Breaks (Had supervised breaks)	5	0.44%	4	0.47%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	0	0.00%	0	0.00%
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	2	0.23%
Z: Read aloud	3	0.26%	1	0.12%
Univ. tool, desig. support, or acc. is in Section 504 plan	1	0.09%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	5	0.44%	7	0.82%
<i>Any Universal tool, desig. support, or accommodation</i>	6	0.53%	7	0.82%
<i>No Universal tool, desig. support, or accommodation</i>	1,133	99.47%	849	99.18%
<b>Target Students Tested</b>	<b>Grade 4</b>	<b>Pct. of Total</b>	<b>Grade 5</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	2	0.28%	1	0.18%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	0	0.00%	0	0.00%
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	2	0.35%
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	2	0.28%	3	0.53%
<i>Any Universal tool, desig. support, or accommodation</i>	2	0.28%	3	0.53%
<i>No Universal tool, desig. support, or accommodation</i>	711	99.72%	568	99.47%

<b>Special Services Summary for RLA, Grades Four and Five</b>				
<b>Nontarget (Optional) Students Tested</b>	<b>Grade 4</b>	<b>Pct. of Total</b>	<b>Grade 5</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	3	0.70%	0	0.00%
K: Breaks (Had supervised breaks)	3	0.70%	3	1.05%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	3	0.70%	1	0.35%
Univ. tool, desig. support, or acc. is in Section 504 plan	1	0.23%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	3	0.70%	4	1.40%
<i>Any</i> Universal tool, desig. support, or accommodation	4	0.94%	4	1.40%
<i>No</i> Universal tool, desig. support, or accommodation	422	99.06%	281	98.60%
<b>Students Not in Special Education</b>	<b>Grade 4</b>	<b>Pct. of Total</b>	<b>Grade 5</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	2	0.18%	0	0.00%
K: Breaks (Had supervised breaks)	2	0.18%	2	0.24%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	1	0.09%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	1	0.09%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	1	0.09%	2	0.24%
<i>Any</i> Universal tool, desig. support, or accommodation	2	0.18%	2	0.24%
<i>No</i> Universal tool, desig. support, or accommodation	1,101	99.82%	823	99.76%

<b>Special Services Summary for RLA, Grades Four and Five</b>				
<b>Students in Special Education</b>	<b>Grade 4</b>	<b>Pct. of Total</b>	<b>Grade 5</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	1	2.86%	0	0.00%
K: Breaks (Had supervised breaks)	3	8.57%	2	6.45%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	0	0.00%	0	0.00%
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	2	6.45%
Z: Read aloud	1	2.86%	1	3.23%
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	3	8.57%	5	16.13%
<i>Any</i> Universal tool, desig. support, or accommodation	3	8.57%	5	16.13%
<i>No</i> Universal tool, desig. support, or accommodation	32	91.43%	26	83.87%
<b>Students in U.S. Schools &lt; 12 Months</b>	<b>Grade 4</b>	<b>Pct. of Total</b>	<b>Grade 5</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	1	0.22%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	0	0.00%	1	0.22%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	1	0.22%
<i>No</i> Universal tool, desig. support, or accommodation	472	100.00%	455	99.78%

<b>Special Services Summary for RLA, Grades Four and Five</b>				
<b>Students in U.S. Schools ≥ 12 Months</b>	<b>Grade 4</b>	<b>Pct. of Total</b>	<b>Grade 5</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	3	0.45%	0	0.00%
K: Breaks (Had supervised breaks)	5	0.75%	3	0.75%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	0	0.00%	0	0.00%
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	2	0.50%
Z: Read aloud	3	0.45%	1	0.25%
Univ. tool, desig. support, or acc. is in Section 504 plan	1	0.15%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	5	0.75%	6	1.50%
<i>Any</i> Universal tool, desig. support, or accommodation	6	0.90%	6	1.50%
<i>No</i> Universal tool, desig. support, or accommodation	661	99.10%	394	98.50%
<b>EL Program: Primary Language Instruction and ELD and/or SDAIE Instruction</b>	<b>Grade 4</b>	<b>Pct. of Total</b>	<b>Grade 5</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	3	0.60%	0	0.00%
K: Breaks (Had supervised breaks)	3	0.60%	3	1.02%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	3	0.60%	1	0.34%
Univ. tool, desig. support, or acc. is in Section 504 plan	1	0.20%	0	0.00%
Univ. tool, desig. support, or acc. is in IEP	3	0.60%	4	1.36%
<i>Any</i> Universal tool, desig. support, or accommodation	4	0.81%	4	1.36%
<i>No</i> Universal tool, desig. support, or accommodation	492	99.19%	291	98.64%

<b>Special Services Summary for RLA, Grades Four and Five</b>				
<b>EL Program: ELD Instruction Only</b>	<b>Grade 4</b>	<b>Pct. of Total</b>	<b>Grade 5</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	24	100.00%	37	100.00%
<b>EL Program: SDAIE Instruction Only</b>	<b>Grade 4</b>	<b>Pct. of Total</b>	<b>Grade 5</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	1	0.65%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	0	0.00%	1	0.65%
<i>Any</i> Universal tool, desig. support, or accommodation	0	0.00%	1	0.65%
<i>No</i> Universal tool, desig. support, or accommodation	137	100.00%	153	99.35%

<b>Special Services Summary for RLA, Grades Four and Five</b>				
<b>EL Program: ELD Instruction and SDAIE Instruction but Not Primary Language Instruction</b>	<b>Grade 4</b>	<b>Pct. of Total</b>	<b>Grade 5</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	2	0.48%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: Unmapped	0	0.00%	0	0.00%
X: Abacus	0	0.00%	0	0.00%
Y: Leave blank	0	0.00%	2	0.65%
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	2	0.48%	2	0.65%
<i>Any</i> Universal tool, desig. support, or accommodation	2	0.48%	2	0.65%
<i>No</i> Universal tool, desig. support, or accommodation	417	99.52%	306	99.35%
<b>EL Program: Other EL Instructional Services</b>	<b>Grade 4</b>	<b>Pct. of Total</b>	<b>Grade 5</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	2	100.00%	4	100.00%

<b>Special Services Summary for RLA, Grades Four and Five</b>				
<b>EL Program: None (EL only)</b>	<b>Grade 4</b>	<b>Pct. of Total</b>	<b>Grade 5</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	36	100.00%	13	100.00%

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

<b>Special Services Summary for RLA, Grades Six and Seven</b>				
<b>All Students Tested</b>	<b>Grade 6</b>	<b>Pct. of Total</b>	<b>Grade 7</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	2	0.38%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	2	0.38%	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	0.19%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	2	0.38%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	521	99.62%	502	100.00%
<b>Target Students Tested</b>	<b>Grade 6</b>	<b>Pct. of Total</b>	<b>Grade 7</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	388	100.00%	416	100.00%

<b>Special Services Summary for RLA, Grades Six and Seven</b>				
<b>Nontarget (Optional) Students Tested</b>	<b>Grade 6</b>	<b>Pct. of Total</b>	<b>Grade 7</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	2	1.48%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	2	1.48%	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	0.74%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	2	1.48%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	133	98.52%	86	100.00%
<b>Students Not in Special Education</b>	<b>Grade 6</b>	<b>Pct. of Total</b>	<b>Grade 7</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	511	100.00%	496	100.00%

<b>Special Services Summary for RLA, Grades Six and Seven</b>				
<b>Students in Special Education</b>	<b>Grade 6</b>	<b>Pct. of Total</b>	<b>Grade 7</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	2	16.67%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	2	16.67%	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	8.33%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	2	16.67%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	10	83.33%	6	100.00%
<b>Students in U.S. Schools &lt; 12 Months</b>	<b>Grade 6</b>	<b>Pct. of Total</b>	<b>Grade 7</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	368	100.00%	392	100.00%

<b>Special Services Summary for RLA, Grades Six and Seven</b>				
<b>Students in U.S. Schools ≥ 12 Months</b>	<b>Grade 6</b>	<b>Pct. of Total</b>	<b>Grade 7</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	2	1.29%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	2	1.29%	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	0.65%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	2	1.29%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	153	98.71%	110	100.00%
<b>EL Program: Primary Language Instruction and ELD and/or SDAIE Instruction</b>	<b>Grade 6</b>	<b>Pct. of Total</b>	<b>Grade 7</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	2	1.22%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	2	1.22%	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	0.61%	0	0.00%
<i>Any</i> Universal tool, desig. support, or accommodation	2	1.22%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	162	98.78%	142	100.00%

<b>Special Services Summary for RLA, Grades Six and Seven</b>				
<b>EL Program: ELD Instruction Only</b>	<b>Grade 6</b>	<b>Pct. of Total</b>	<b>Grade 7</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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%	11	100.00%
<b>EL Program: SDAIE Instruction Only</b>	<b>Grade 6</b>	<b>Pct. of Total</b>	<b>Grade 7</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	25	100.00%	8	100.00%

<b>Special Services Summary for RLA, Grades Six and Seven</b>				
<b>EL Program: ELD Instruction and SDAIE Instruction but Not Primary Language Instruction</b>	<b>Grade 6</b>	<b>Pct. of Total</b>	<b>Grade 7</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	304	100.00%	312	100.00%
<b>EL Program: Other EL Instructional Services</b>	<b>Grade 6</b>	<b>Pct. of Total</b>	<b>Grade 7</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	3	100.00%	3	100.00%

<b>Special Services Summary for RLA, Grades Six and Seven</b>				
<b>EL Program: None (EL only)</b>	<b>Grade 6</b>	<b>Pct. of Total</b>	<b>Grade 7</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	11	100.00%	9	100.00%

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

<b>Special Services Summary for RLA, Grades Eight and Nine</b>				
<b>All Students Tested</b>	<b>Grade 8</b>	<b>Pct. of Total</b>	<b>Grade 9</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	1	0.21%	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	1	0.21%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	1	0.21%	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	0.21%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	481	99.79%	1,049	100.00%
<b>Target Students Tested</b>	<b>Grade 8</b>	<b>Pct. of Total</b>	<b>Grade 9</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	1	0.22%	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	1	0.22%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	1	0.22%	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	0.22%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	449	99.78%	975	100.00%

<b>Special Services Summary for RLA, Grades Eight and Nine</b>				
<b>Nontarget (Optional) Students Tested</b>	<b>Grade 8</b>	<b>Pct. of Total</b>	<b>Grade 9</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	32	100.00%	74	100.00%
<b>Students Not in Special Education</b>	<b>Grade 8</b>	<b>Pct. of Total</b>	<b>Grade 9</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	1	0.21%	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	1	0.21%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	1	0.21%	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	0.21%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	475	99.79%	1,047	100.00%

<b>Special Services Summary for RLA, Grades Eight and Nine</b>				
<b>Students in Special Education</b>	<b>Grade 8</b>	<b>Pct. of Total</b>	<b>Grade 9</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	6	100.00%	2	100.00%
<b>Students in U.S. Schools &lt; 12 Months</b>	<b>Grade 8</b>	<b>Pct. of Total</b>	<b>Grade 9</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	1	0.24%	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	1	0.24%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	1	0.24%	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	0.24%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	420	99.76%	932	100.00%

<b>Special Services Summary for RLA, Grades Eight and Nine</b>				
<b>Students in U.S. Schools ≥ 12 Months</b>	<b>Grade 8</b>	<b>Pct. of Total</b>	<b>Grade 9</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	61	100.00%	117	100.00%
<b>EL Program: Primary Language Instruction and ELD and/or SDAIE Instruction</b>	<b>Grade 8</b>	<b>Pct. of Total</b>	<b>Grade 9</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	82	100.00%	250	100.00%

<b>Special Services Summary for RLA, Grades Eight and Nine</b>				
<b>EL Program: ELD Instruction Only</b>	<b>Grade 8</b>	<b>Pct. of Total</b>	<b>Grade 9</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	10	100.00%	11	100.00%
<b>EL Program: SDAIE Instruction Only</b>	<b>Grade 8</b>	<b>Pct. of Total</b>	<b>Grade 9</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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%	9	100.00%

<b>Special Services Summary for RLA, Grades Eight and Nine</b>				
<b>EL Program: ELD Instruction and SDAIE Instruction but Not Primary Language Instruction</b>	<b>Grade 8</b>	<b>Pct. of Total</b>	<b>Grade 9</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	1	0.28%	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	1	0.28%	0	0.00%
Univ. tool, desig. support, or acc. is in Section 504 plan	1	0.28%	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	0.28%	0	0.00%
<i>No</i> Universal tool, desig. support, or accommodation	350	99.72%	678	100.00%
<b>EL Program: Other EL Instructional Services</b>	<b>Grade 8</b>	<b>Pct. of Total</b>	<b>Grade 9</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	0	0.00%	11	100.00%

<b>Special Services Summary for RLA, Grades Eight and Nine</b>				
<b>EL Program: None (EL only)</b>	<b>Grade 8</b>	<b>Pct. of Total</b>	<b>Grade 9</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	13	100.00%	22	100.00%

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

<b>Special Services Summary for RLA, Grades Ten and Eleven</b>				
<b>All Students Tested</b>	<b>Grade 10</b>	<b>Pct. of Total</b>	<b>Grade 11</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	410	100.00%	198	100.00%
<b>Target Students Tested</b>	<b>Grade 10</b>	<b>Pct. of Total</b>	<b>Grade 11</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	350	100.00%	172	100.00%

<b>Special Services Summary for RLA, Grades Ten and Eleven</b>				
<b>Nontarget (Optional) Students Tested</b>	<b>Grade 10</b>	<b>Pct. of Total</b>	<b>Grade 11</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	60	100.00%	26	100.00%
<b>Students Not in Special Education</b>	<b>Grade 10</b>	<b>Pct. of Total</b>	<b>Grade 11</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	409	100.00%	197	100.00%

<b>Special Services Summary for RLA, Grades Ten and Eleven</b>				
<b>Students in Special Education</b>	<b>Grade 10</b>	<b>Pct. of Total</b>	<b>Grade 11</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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%	1	100.00%
<b>Students in U.S. Schools &lt; 12 Months</b>	<b>Grade 10</b>	<b>Pct. of Total</b>	<b>Grade 11</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	340	100.00%	161	100.00%

<b>Special Services Summary for RLA, Grades Ten and Eleven</b>				
<b>Students in U.S. Schools ≥ 12 Months</b>	<b>Grade 10</b>	<b>Pct. of Total</b>	<b>Grade 11</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	70	100.00%	37	100.00%
<b>EL Program: Primary Language Instruction and ELD and/or SDAIE Instruction</b>	<b>Grade 10</b>	<b>Pct. of Total</b>	<b>Grade 11</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	151	100.00%	56	100.00%

<b>Special Services Summary for RLA, Grades Ten and Eleven</b>				
<b>EL Program: ELD Instruction Only</b>	<b>Grade 10</b>	<b>Pct. of Total</b>	<b>Grade 11</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	10	100.00%	6	100.00%
<b>EL Program: SDAIE Instruction Only</b>	<b>Grade 10</b>	<b>Pct. of Total</b>	<b>Grade 11</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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%	5	100.00%

<b>Special Services Summary for RLA, Grades Ten and Eleven</b>				
<b>EL Program: ELD Instruction and SDAIE Instruction but Not Primary Language Instruction</b>	<b>Grade 10</b>	<b>Pct. of Total</b>	<b>Grade 11</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	209	100.00%	114	100.00%
<b>EL Program: Other EL Instructional Services</b>	<b>Grade 10</b>	<b>Pct. of Total</b>	<b>Grade 11</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	2	100.00%	1	100.00%

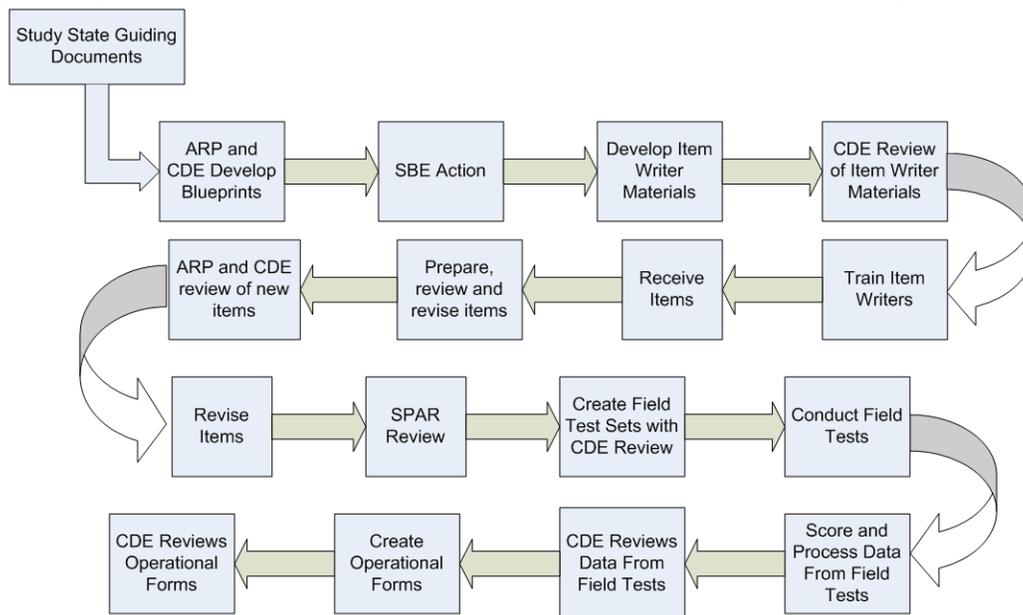
<b>Special Services Summary for RLA, Grades Ten and Eleven</b>				
<b>EL Program: None (EL only)</b>	<b>Grade 10</b>	<b>Pct. of Total</b>	<b>Grade 11</b>	<b>Pct. of Total</b>
B: Marked in test booklet	0	0.00%	0	0.00%
C: Scribe	0	0.00%	0	0.00%
F: Alternative response options	0	0.00%	0	0.00%
G: Braille	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)	0	0.00%	0	0.00%
K: Breaks (Had supervised breaks)	0	0.00%	0	0.00%
L: Administered at most beneficial time of day	0	0.00%	0	0.00%
M: Separate setting	0	0.00%	0	0.00%
N: Dictionary, thesaurus	0	0.00%	0	0.00%
O: American Sign Language	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%
W: 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	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	16	100.00%	6	100.00%

## Chapter 3: Item Development

The intact or modified intact test forms from previous test administrations from different years were used during the 2014 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 forms.

The STS items were developed to measure California’s content standards and designed to conform to principles of item writing defined by ETS (ETS, 2002). Each STS item on the intact or modified intact forms used in 2014 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 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 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. The specifications included the following:

- A full statement of each academic content standard, as defined by the 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, if 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 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 STS administration, field-test items were repeated as a part of the intact or modified intact forms.

## **Selection of Item Writers**

### **Criteria for Selecting Item Writers**

The items for each STS were developed by individual item writers with a thorough understanding of the California content standards. 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. 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 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 items and test forms—currently being reused—at the time the items and forms were developed. See Table 8.4 on page 145 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 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 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. The ARPs are advisory panels to the CDE and ETS and provided guidance on matters related to

item development for the STS. 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. 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; and
- Bilingual and biliterate in Spanish and English.

School administrators, 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; 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 Items**

ETS content-area assessment specialists facilitated the STS 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
- Overview of the STS’s test design specifications and blueprints
- Analysis of the STS’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
- 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 *EC* 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 or modified intact forms are being used with the field-test items intact for the 2014 CAASPP administration, data were not analyzed for current field-test items.

### **Stand-alone Field Testing**

For each new STS 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 administration, the original field-test items remained in their original positions in the

intact or modified intact forms. Data were not analyzed for field-test items. The numbers of embedded field-test items for the STS are not presented in this report because for the 2014 administration, field-test items were repeated as a part of the intact or modified intact forms and there was no new item development.

### **Allocation of Students to Forms**

The test forms for a given STS 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 IRT 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 (CSTs, CMA, CAPA, and STS). 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 STS were constructed to measure students' performance relative to California's content standards approved by the SBE. They were also constructed to meet professional standards for validity and reliability. For each STS, 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 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 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, six items are field-test items. For more details on the distribution of items, see Appendix 2.A—STS Items and Time Chart on page 18.

### Rules for Item Selection

#### Test Blueprint

All test items on STS forms were selected to conform to the SBE-approved California content standards and test blueprints. The content blueprints for the STS can be found on the 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 items are grouped into subcontent areas (reporting clusters). For each STS reporting cluster, the percentage of questions correctly answered is reported on a student's score report. A list of the STS 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

The intact or modified intact test forms from different years of administration were used during the 2014 administration. Prior to the 2014 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 58, contains further information about this process.

### **Replacement Items**

For the 2014 STS administration, both the grade six and grade eight forms included replacement items—one replacement item in grade six and five replacement items in grade eight.

Before replacing items, ETS considered whether the form with an item requiring replacement could be exchanged with a form from a different previous administration. If there were no appropriate forms available, ETS sought to replace the items with previously used or field-tested items that best met the psychometric criteria and statistical targets described in the next sections. As the STS would be pre-equated, ETS assessment development staff also considered minimizing the position difference between the replacement items and the exposed items in order to reduce the context effect.

### **Psychometric Criteria**

The three goals of STS 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. These rules are referred to as statistical targets.

Two types of assembly targets were developed for each STS: 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 forms for the intact or modified intact forms used in the 2014 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 the total test are presented in Figure 4.A.1, on page 74. These curves present the target TIF and the projected TIF for the total test.

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 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 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 Test Assembly**

Content Area	STS *	Target Mean $b$	Target SD $b$	Min $p$ -value	Max $p$ -value	Mean Point Biserial	Min Point Biserial
	2	-0.44	0.91	0.20	0.95	> 0.37	0.14
	3	-0.46	0.88	0.20	0.95	> 0.37	0.14
	4	-0.45	0.71	0.20	0.95	> 0.37	0.14
	5	-0.24	0.65	0.20	0.95	> 0.37	0.14
Reading/ Language Arts	6	-0.23	0.72	0.20	0.95	> 0.37	0.14
	7	-0.21	0.80	0.20	0.95	> 0.37	0.14
	8	-0.20	0.74	0.20	0.95	> 0.37	0.14
	9	-0.21	0.77	0.20	0.95	> 0.37	0.14
	10	-0.21	0.79	0.20	0.95	> 0.37	0.14
	11	-0.21	0.70	0.20	0.95	> 0.37	0.14

\* Numbers indicate grade-level tests.

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 76, present the target and projected information curves for the reporting clusters in the administered tests.

## Projected Psychometric Properties of the Assembled Tests

Prior to the 2014 administration, ETS 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 73 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 STS administration.

Table 4.A.2 on page 73 shows the mean observed statistics of the items on each STS based on the item-level statistics from the year the form was previously administered. See Table 8.4 on page 145 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 2014 STS Projected Technical Characteristics**

Content Area	STS *	Number of Op. Items	Mean Raw Score	Std. Dev. of Raw Scores	Reliability
Reading/ Language Arts	2	65	39.59	12.85	0.93
	3	65	35.87	11.89	0.91
	4	75	43.10	14.38	0.93
	5	75	36.52	13.16	0.91
	6	75	38.94	13.39	0.91
	7	75	39.30	12.74	0.91
	8	75	39.75	12.98	0.91
	9	75	40.76	11.25	0.88
	10	75	41.07	12.20	0.90
	11	75	37.82	10.71	0.86

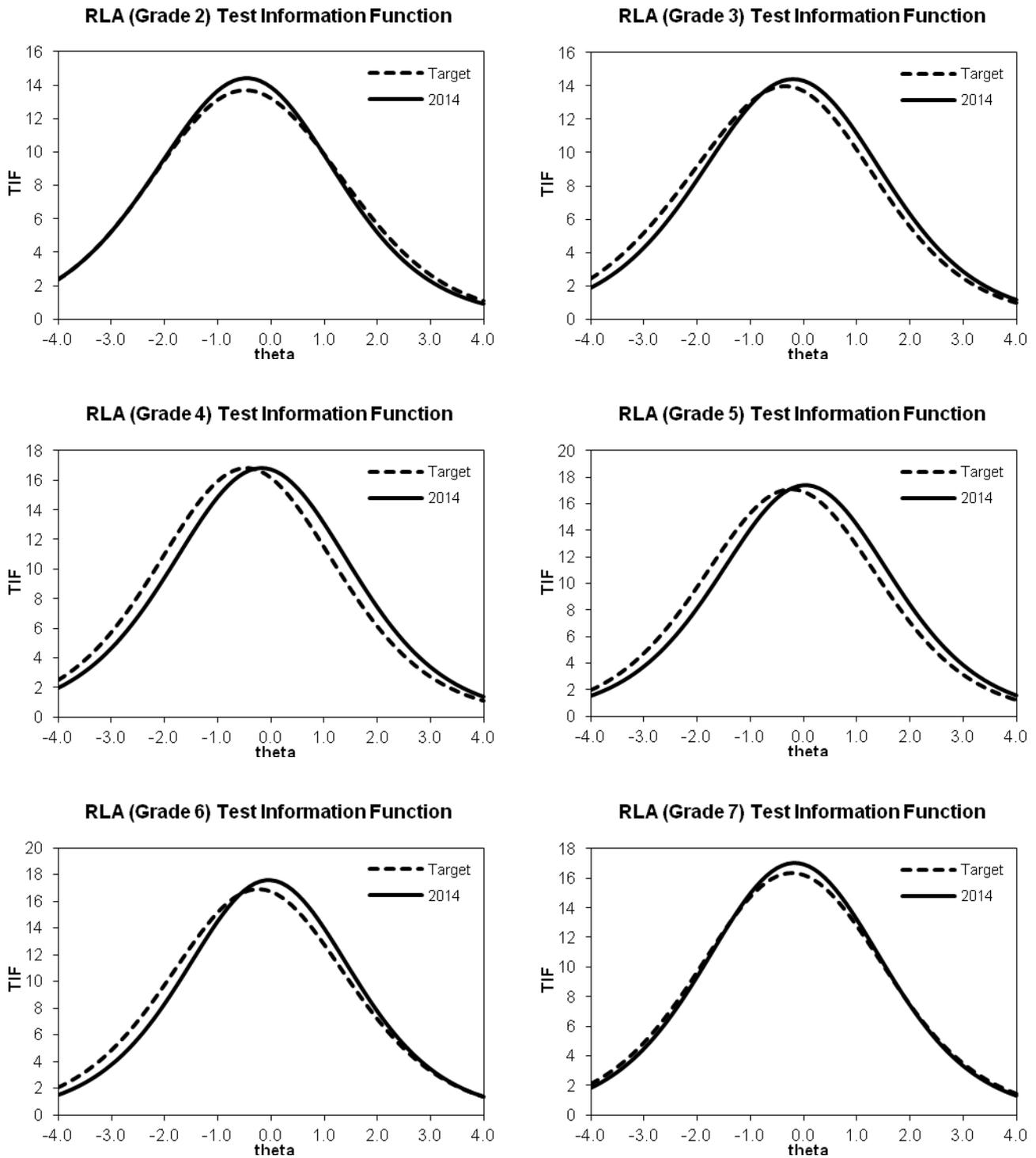
\* Numbers indicate grade-level tests.

**Table 4.A.2 Summary of 2014 STS Projected Statistical Attributes**

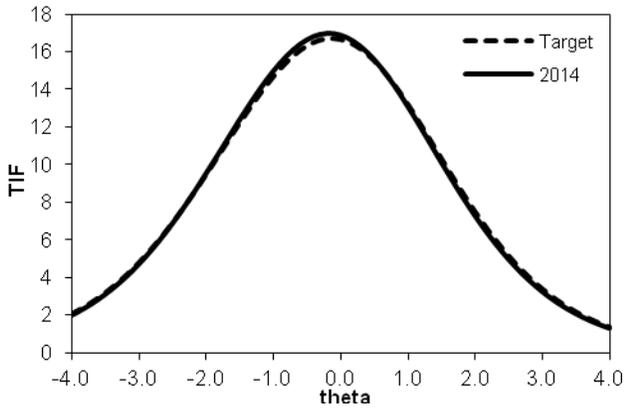
Content Area	STS *	Mean b	SD b	Mean $p$ -value	Min $p$ -value	Max $p$ -value	Mean Point Biserial	Min Point Biserial
Reading/ Language Arts	2	-0.51	0.76	0.61	0.28	0.92	0.42	0.20
	3	-0.24	0.76	0.57	0.30	0.92	0.38	0.20
	4	-0.19	0.70	0.57	0.28	0.89	0.41	0.19
	5	0.02	0.58	0.50	0.25	0.83	0.37	0.20
	6	-0.04	0.46	0.51	0.30	0.75	0.36	0.15
	7	-0.17	0.65	0.56	0.30	0.79	0.37	0.17
	8	-0.21	0.67	0.54	0.26	0.83	0.36	0.19
	9	-0.18	0.73	0.55	0.26	0.84	0.35	0.17
	10	-0.24	0.74	0.56	0.28	0.88	0.34	0.20
	11	-0.19	0.69	0.56	0.22	0.90	0.34	0.08

\* Numbers indicate grade-level tests.

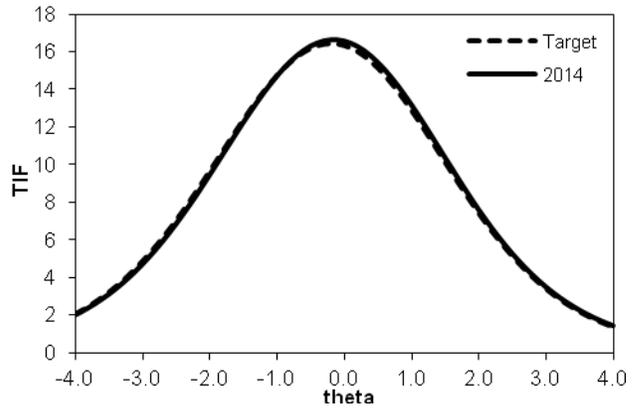
**Figure 4.A.1 Plots of Target Information Function and Projected Information for Total Test and Linking Set for RLA**



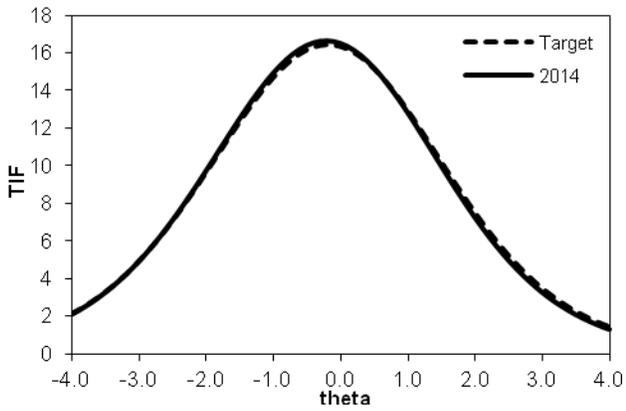
**RLA (Grade 8) Test Information Function**



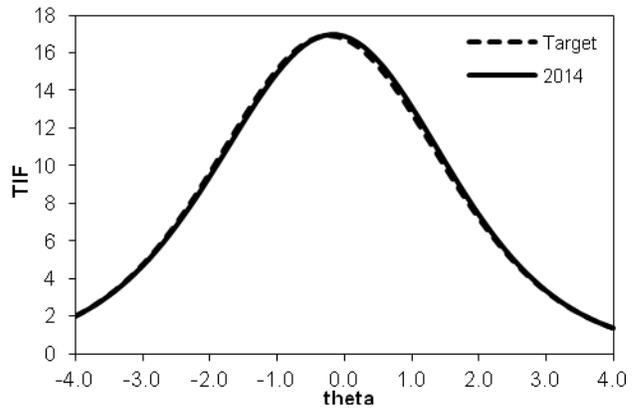
**RLA (Grade 9) Test Information Function**



**RLA (Grade 10) Test Information Function**

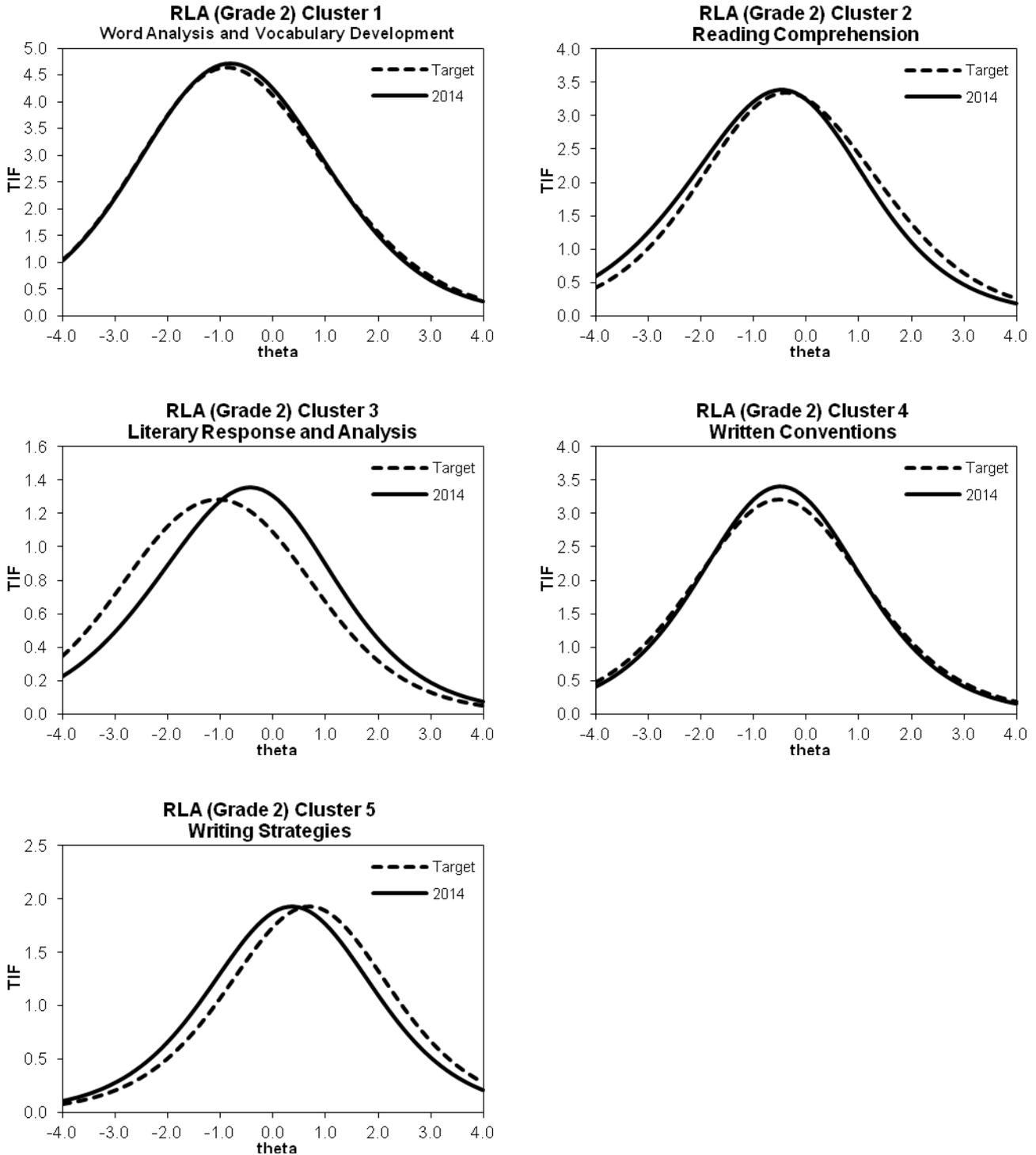


**RLA (Grade 11) Test Information Function**

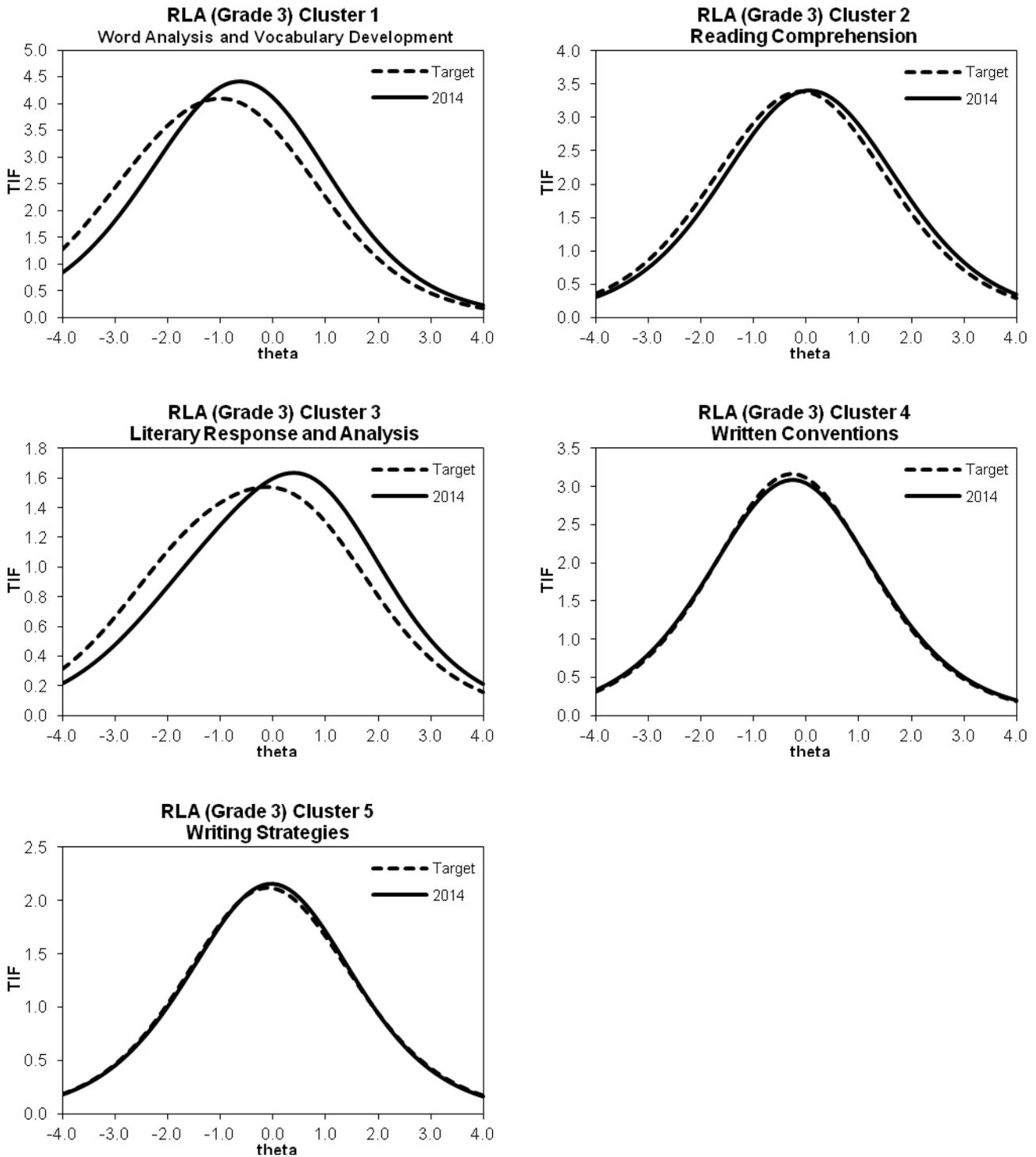


## Appendix 4.B—Cluster Targets

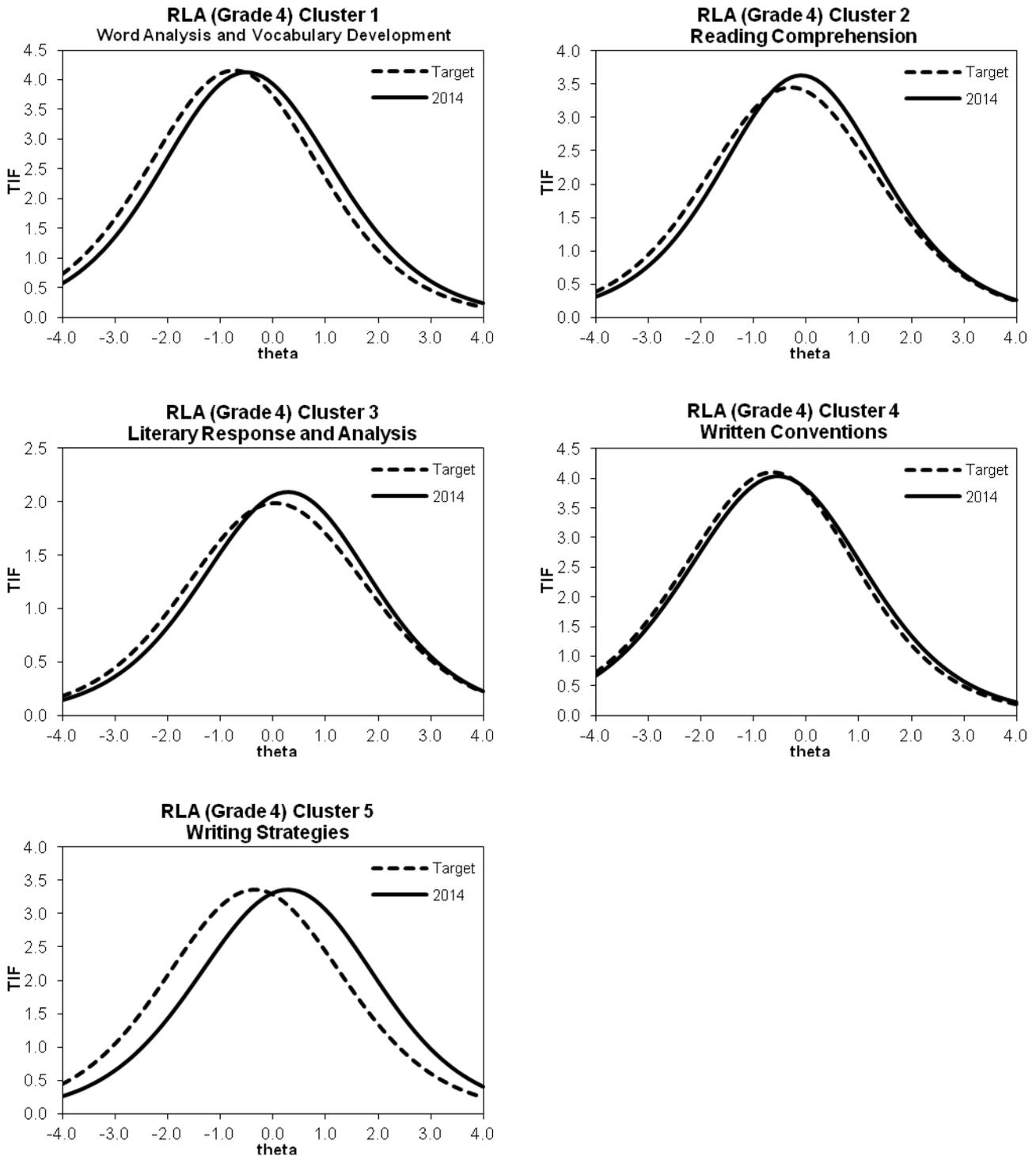
Figure 4.B.1 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Two



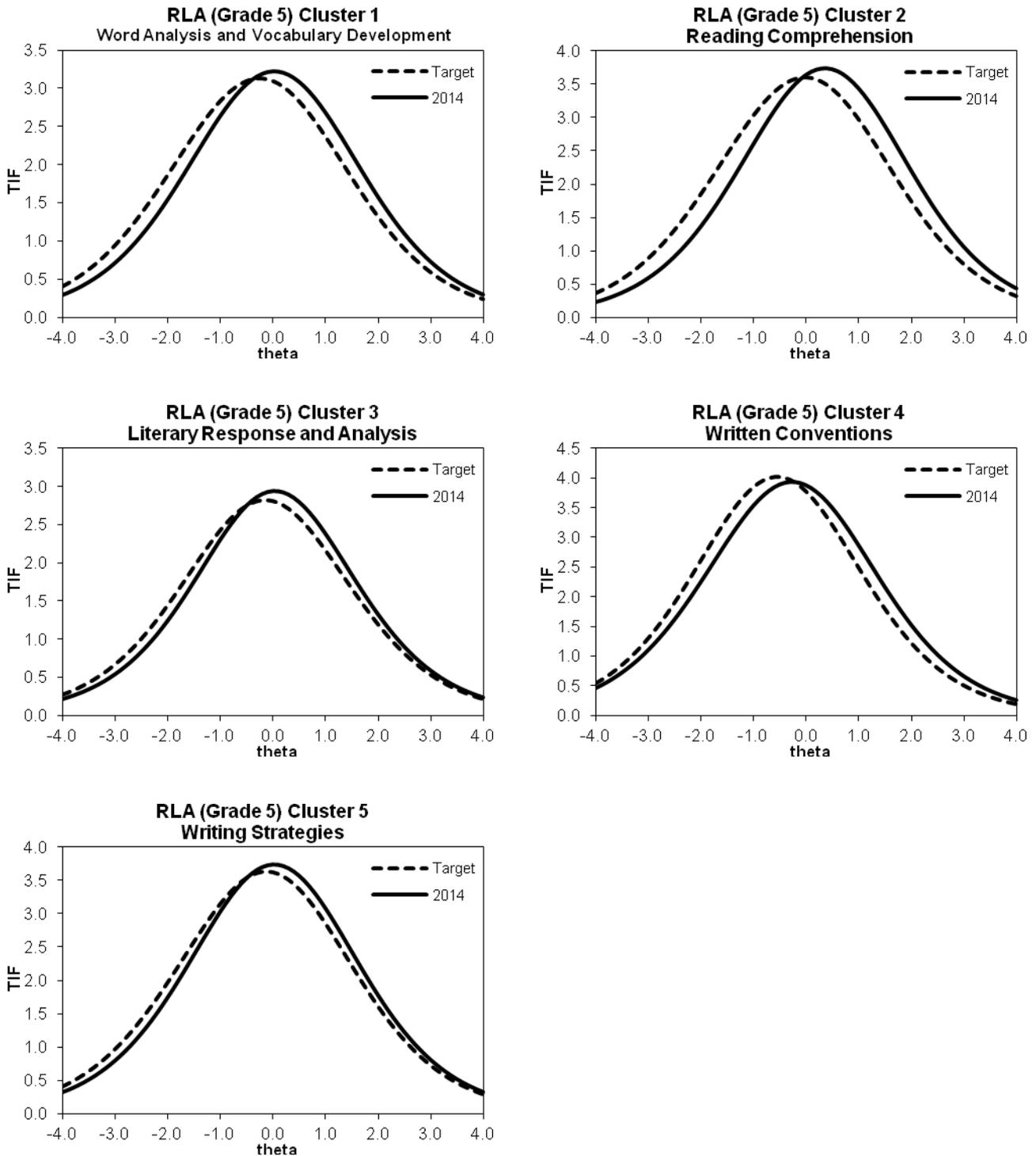
**Figure 4.B.2 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Three**



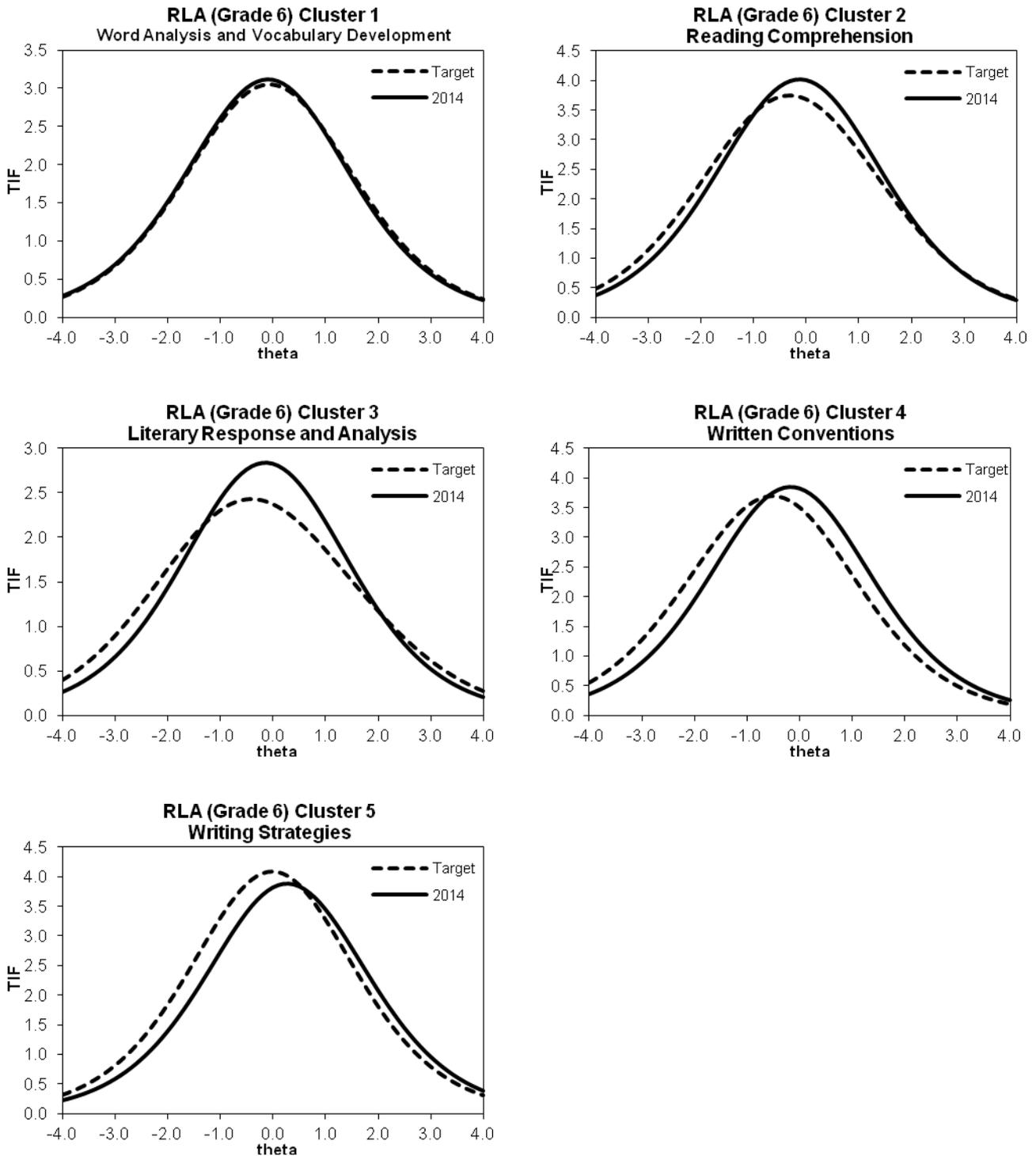
**Figure 4.B.3 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Four**



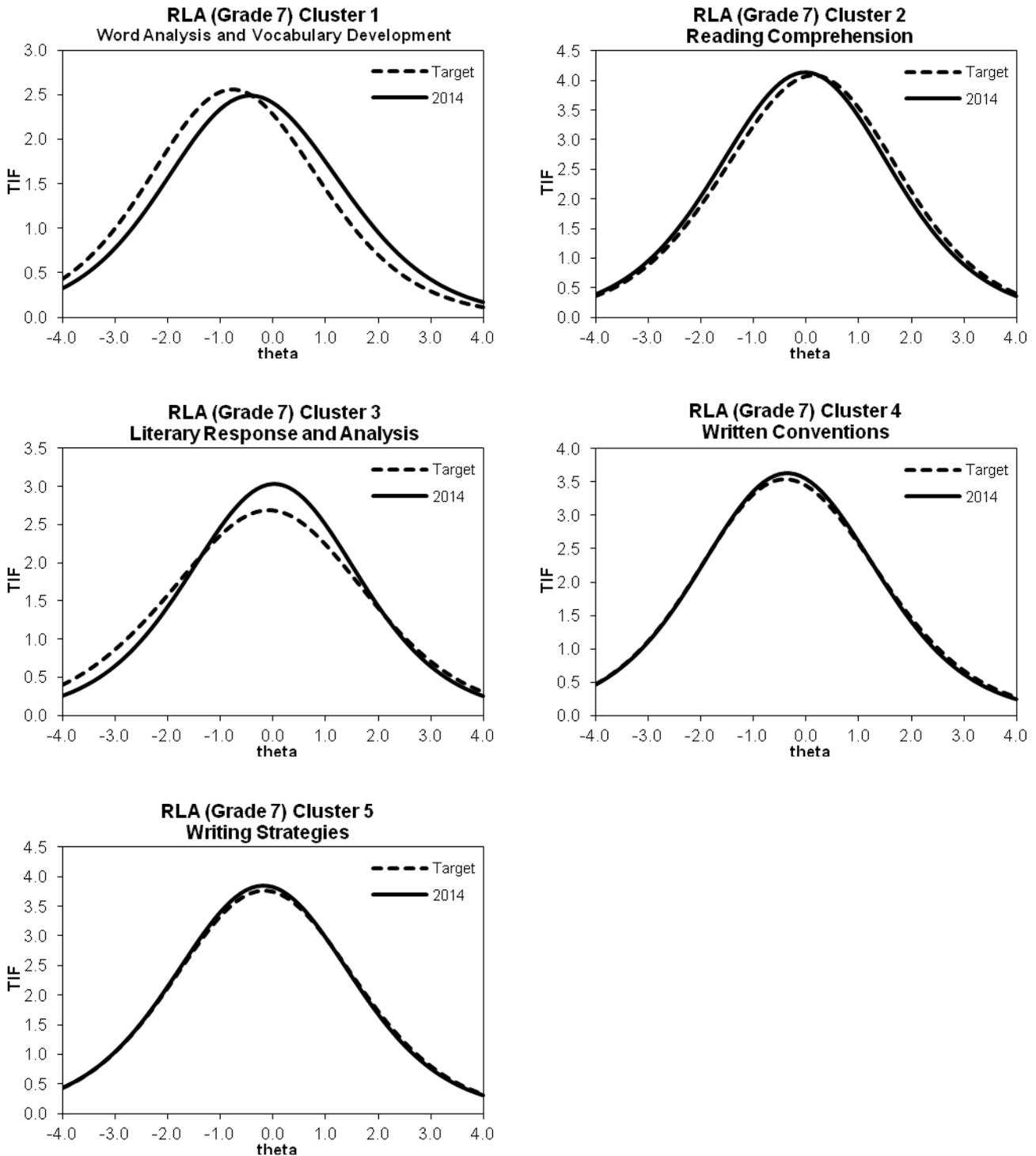
**Figure 4.B.4 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Five**



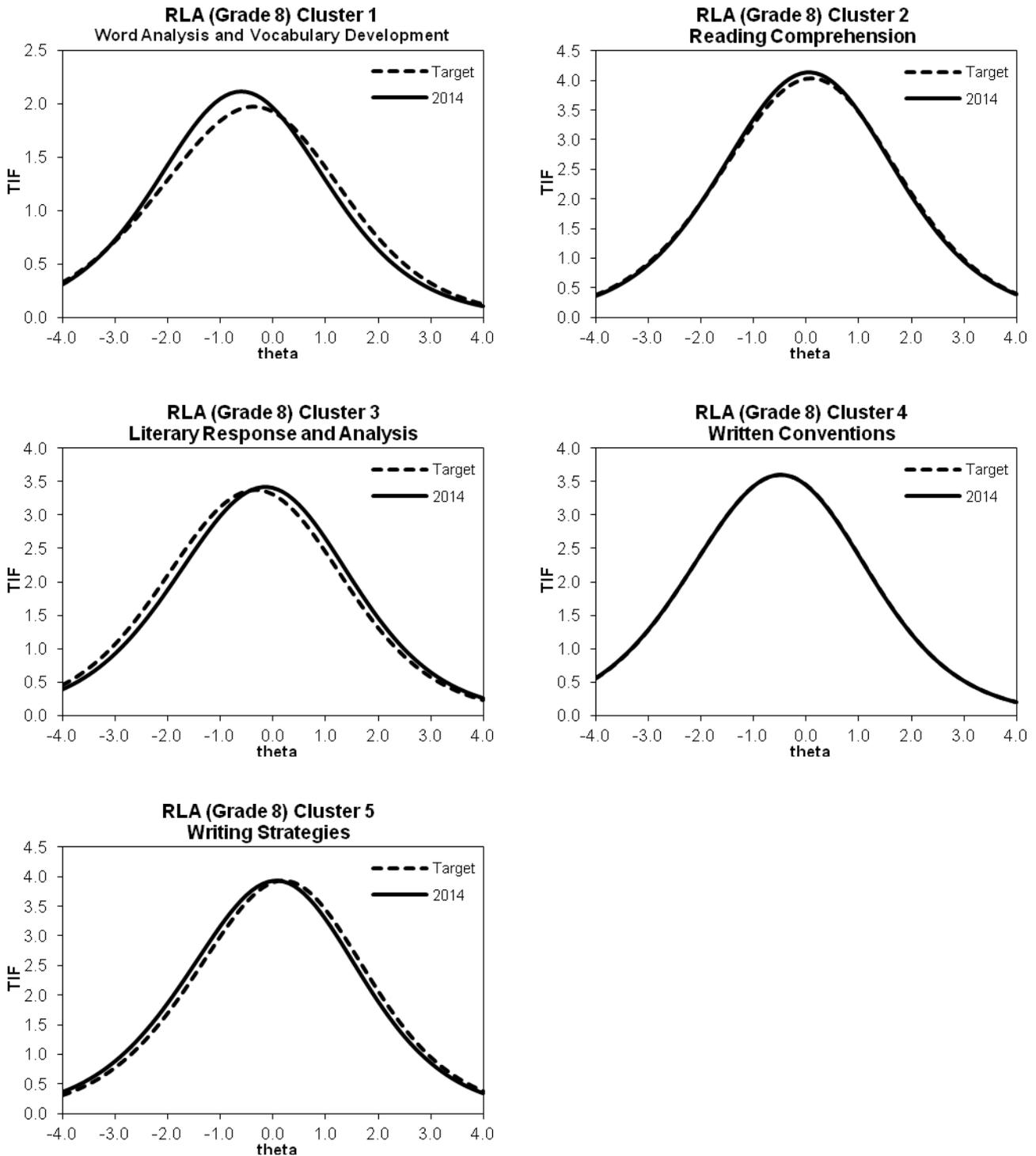
**Figure 4.B.5 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Six**



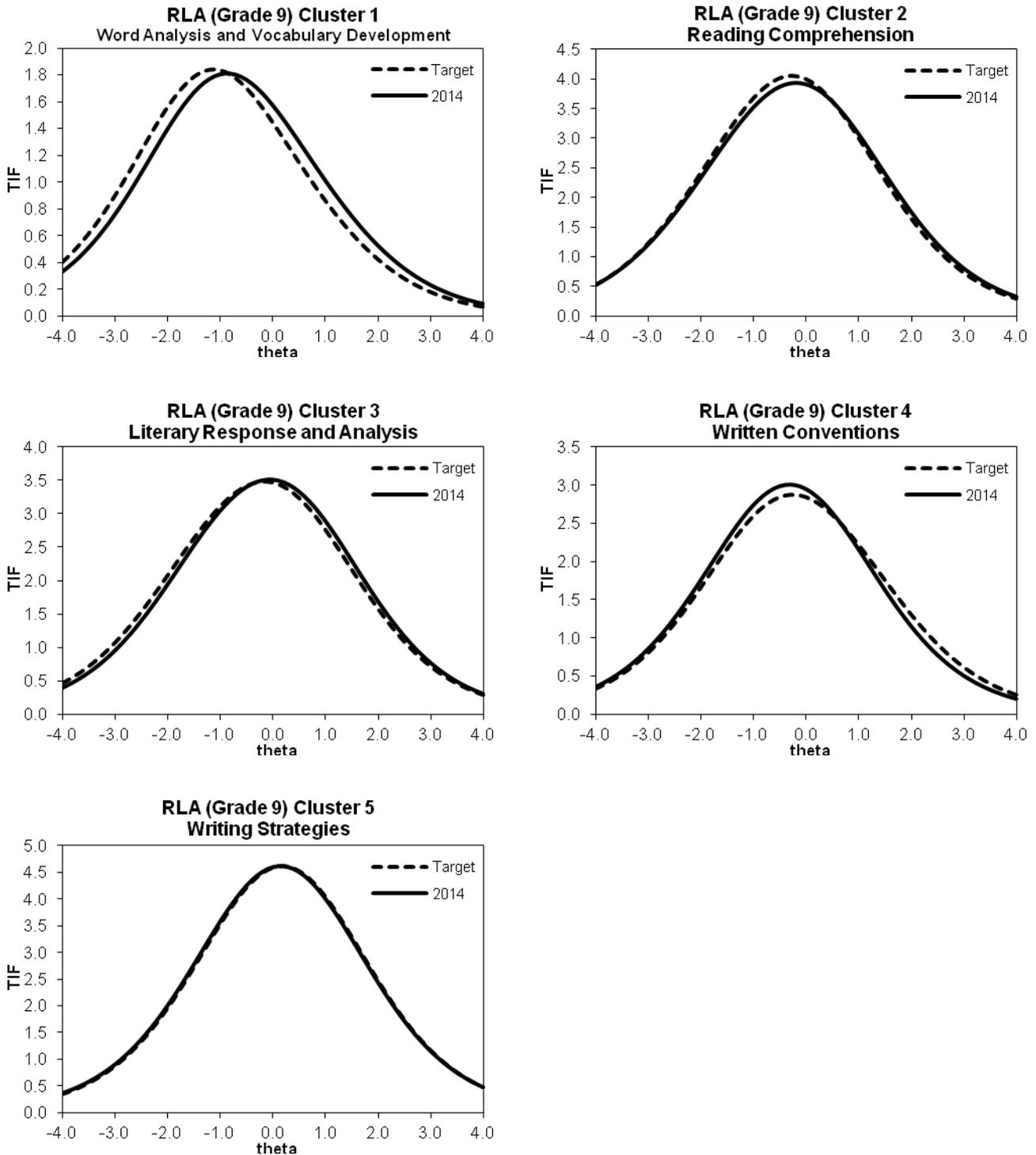
**Figure 4.B.6 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Seven**



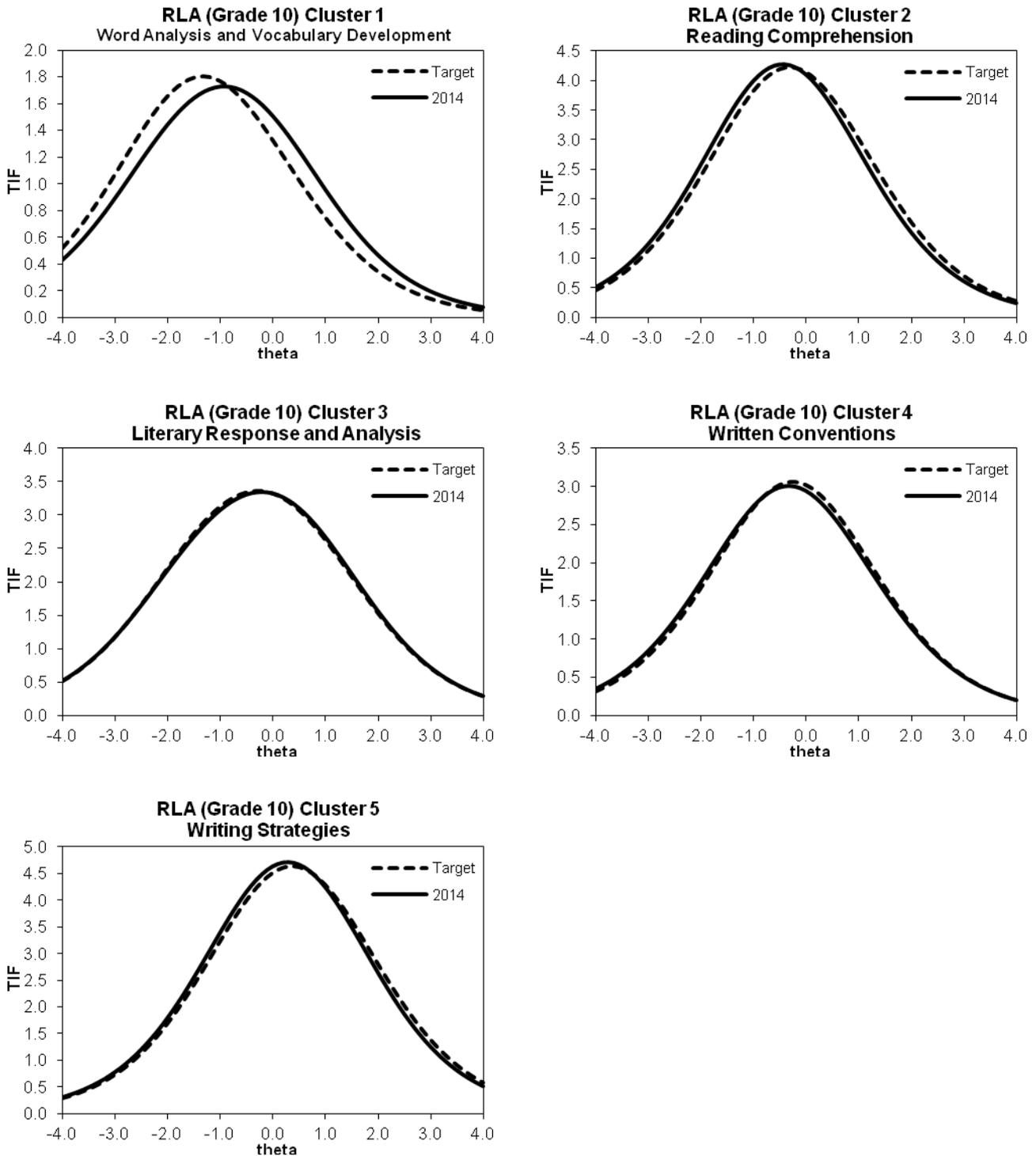
**Figure 4.B.7 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Eight**



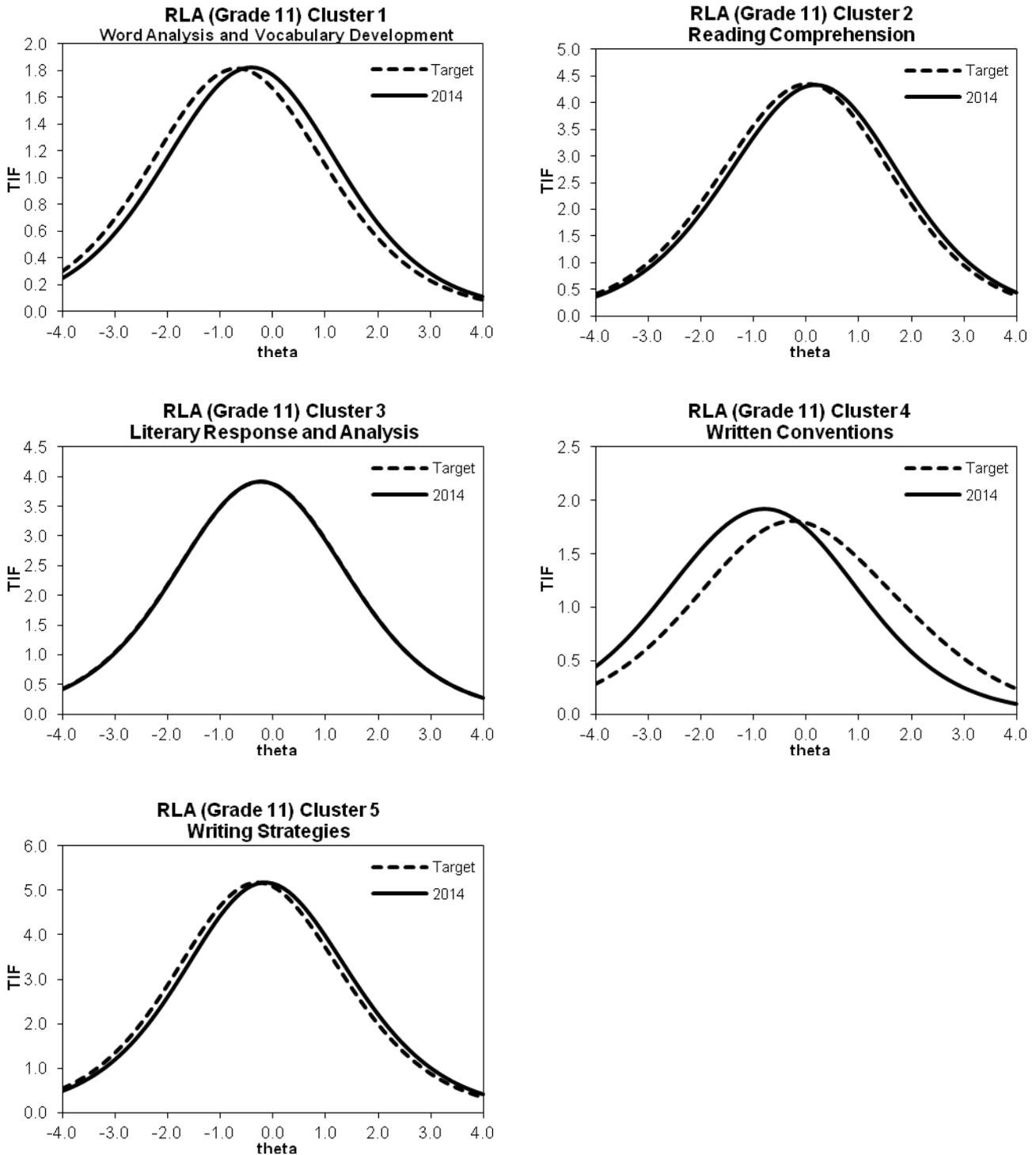
**Figure 4.B.8 Plots of Target Information Functions and Projected Information for Clusters for RLA, Grade Nine**



**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 CAASPP System are secure documents. For the STS administration, every person having access to testing 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), 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

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, ARP meetings were not held in 2014 because there was no new item development for the 2014 STS forms. However, before the 2014 administration,

ETS facilitated ARP meetings every year to review all newly developed STS 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 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 Management System (CDE, 2014a) remain protected by the ETS firewall software at all times. Due to the sensitive nature of the student

information processed by the Test Management System, 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 Pearson and other authorized locations. Pearson distributes the booklets to LEAs in securely packaged boxes.

## Test Administration

Pearson receives testing materials from printers, packages them, and sends them to LEAs. After testing, the LEAs return materials to Pearson for scoring. During these events, Pearson takes extraordinary measures to protect the testing materials. Pearson's customized Oracle business applications verify that inventory controls are in place, from materials receipt to packaging. The reputable carriers used by Pearson 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 Pearson. Pearson 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 Pearson identification badges at all times in Pearson 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

Pearson provides overall security for assessment materials through its limited-access facilities and through its secure data processing capabilities. Pearson 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 Pearson 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 Pearson 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, Pearson utilizes Random Access Control Facility (RACF) to limit and control access to all data files (test and production), source code, object code, databases, and tables. RACF controls 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, Pearson uses ChangeMan, a mainframe configuration management tool, to control versions of the software and data files. ChangeMan provides another level of security, combined with RACF, to place the correct tested version of code into production. Unapproved changes are not implemented without prior review and approval.

## **Transfer of Scores via Secure Data Exchange**

After scoring is completed, Pearson sends scored data files to ETS using secure data exchange procedures. ETS and Pearson have implemented procedures and systems to provide efficient coordination of secure data exchange. This includes the established SFTP site that is used for secure data transfers between ETS and Pearson. These well-established procedures provide timely, efficient, and secure transfer of data. Access to the CAASPP data files is limited to appropriate personnel with direct project responsibilities.

## **Statistical Analysis**

The Information Technology (IT) area at ETS retrieves the Pearson data files from the SFTP site and loads them into a database. The Data Quality Services (DQS) area 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:

- Paper reports, some with individual student results and others with summary results
- A file of individual student results—available for download through the electronic reporting function of the Test Management System’s Quick-turnaround Reporting (QTR) module—that shows students’ scale scores and performance levels
- 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 ESEA 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 (NSLP), and so forth (CDE, 2014b). 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 Data 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 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, as described in this section.

### Test Administrators

The STS 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 LEA and Test Site Coordinator Manual* (CDE, 2014c). 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. (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, 2014c)

### **Test Examiner**

The STS are administered by test examiners who may be assisted by test proctors and scribes. A test examiner 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 examiners must follow the directions in the *California Standards-based Tests in Spanish Directions for Administration (DFA)* (CDE, 2014d) exactly.

### **Test Proctor**

A test proctor is an employee of an LEA or a person, assigned by an NPS to implement the 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**

STS DFAs are manuals used by test examiners to administer the STS to students (CDE, 2014d). Test examiners must follow all directions and guidelines and read, word-for-word, the instructions to students in "SAY" boxes to ensure test standardization.

## **CAASPP LEA and Test Site Coordinator Manual**

Test administration procedures are to be followed exactly so that all students have an equal opportunity to demonstrate their academic achievement. The *CAASPP LEA and Test Site Coordinator 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, 2014c). However, the manual is

not intended as a substitute for the *CCR, Title 5, Education (5 CCR)*, or to detail all of the coordinator’s responsibilities.

### Test Management System Manuals

The Test Management System is a series of secure, Web-based modules that allow LEA CAASPP Coordinators to set up test administrations, order materials, and submit and correct student Pre-ID data. Every module has its own user manual with detailed instructions on how to use the Test Management System. The modules of the Test Management System are as follows:

- **Test Administration Setup**—This module allows LEAs to determine and calculate dates for scheduling test administrations for LEAs, to verify contact information for those LEAs, and to update the LEA’s shipping information. (CDE, 2014e)
- **Order Management**—This module allows LEAs to enter quantities of testing materials for schools. Its manual includes guidelines for determining which materials to order. (CDE, 2014f)
- **Pre-ID**—This module allows LEAs to enter or upload student information, including some demographics, and identify the test(s) the student will take. This information is printed on student test booklets or answer documents or on labels that can be affixed to test booklets or answer documents. Its manual includes the CDE’s Pre-ID layout. (CDE, 2014b)
- **Extended Pre-ID Data Corrections**—This module allows LEAs to correct the data that were submitted during Pre-ID prior to the last day of the LEA’s selected testing window. (CDE, 2014b)

### 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 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, depending on how the LEA CAASPP Coordinator ordered the materials. 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 ELs. 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 under standard conditions. However, some students with disabilities may need assistance when taking the STS. 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, 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 is to enable students to take the STS, not to give them an advantage over other students or to inflate their scores artificially. Scores for students tested with non-embedded accessibility supports are counted as far below basic for aggregate reporting; universal tools, designated supports, or accommodations do not result in changes to students' scores.

## 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 LEA and Test Site Coordinator Manual* (CDE, 2014c).

## 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 high-risk social media security breaches associated with the STS in 2014 that required any item to be withheld from scoring.

## Testing Improprieties

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, 2014d) and the *CAASPP LEA and Test Site Coordinator Manual* (CDE, 2014c). 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. (2014a). *2014 Test Management System*. Sacramento, CA. <http://caaspp.org/administration/tms/>
- California Department of Education. (2014b). *2014 CAASPP Pre-ID and Extended Pre-ID Data Corrections instructions manual*. Sacramento, CA. Retrieved from [http://www.startest.org/pdfs/calif-tac.pre-id\\_xdc\\_manual.2014.pdf](http://www.startest.org/pdfs/calif-tac.pre-id_xdc_manual.2014.pdf)
- California Department of Education. (2014c). *2014 CAASPP LEA and test site coordinator manual*. Sacramento, CA. Retrieved from [http://caaspp.org/rsc/pdfs/CAASPP.coord\\_man.2014.pdf](http://caaspp.org/rsc/pdfs/CAASPP.coord_man.2014.pdf)
- California Department of Education. (2014d). *2014 California Standards-based Tests in Spanish directions for administration*. Sacramento, CA. Retrieved from [http://caaspp.org/rsc/pdfs/STS.grade-5\\_dfa.2014.pdf](http://caaspp.org/rsc/pdfs/STS.grade-5_dfa.2014.pdf)
- California Department of Education. (2014e). *2014 CAASPP Test Administration Setup manual*. Sacramento, CA. Retrieved from [http://www.startest.org/pdfs/STAR.test\\_admin\\_setup.2014.pdf](http://www.startest.org/pdfs/STAR.test_admin_setup.2014.pdf)
- California Department of Education. (2014f). *2014 CAASPP Order Management manual*. Sacramento, CA. Retrieved from [http://www.startest.org/pdfs/calif-tac.order\\_mgmt.2014.pdf](http://www.startest.org/pdfs/calif-tac.order_mgmt.2014.pdf)

# Chapter 6: Performance Standards

---

## Background

The 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 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 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 EOC 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.

The state target is to have all students achieve the proficient or advanced level by 2014. Schools and LEAs are expected to provide additional assistance to students scoring at or below the basic level.

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. 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 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 and recruited from diverse geographic regions and from different gender and major racial/ethnic subgroups to be

representative of the educators of the state’s STS-eligible students (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, 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 tests.

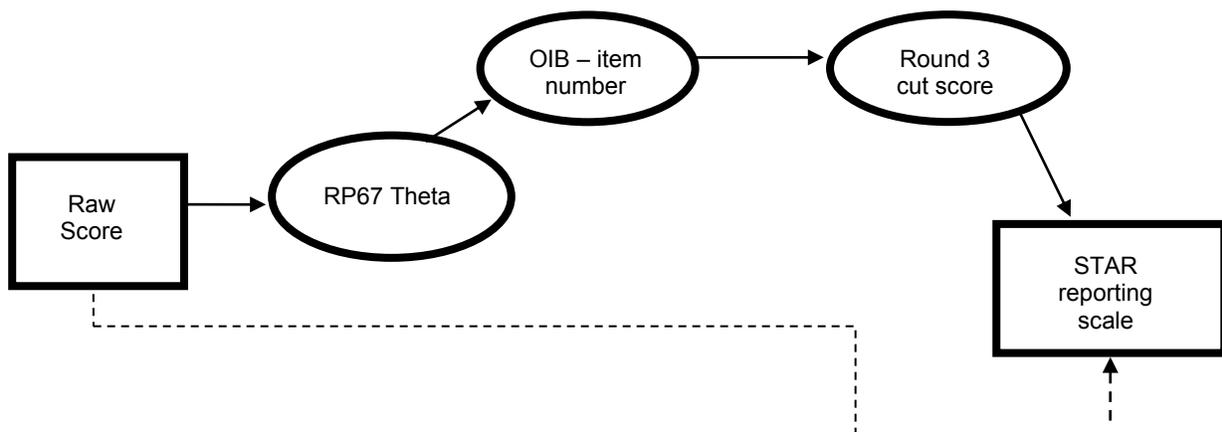
The Bookmark method is an item-mapping procedure in which panelists consider content covered by items in a specially constructed book where items are ordered from easiest to hardest based on operational student performance data from a previous test administration. The “item map,” which accompanies the 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 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* ( $\theta$ ) 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

$\theta_{\text{basic}}$  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 7.4 on page 106 in Chapter 7 presents the percentages of examinees meeting each performance level in 2014.

## 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

---

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 STS 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 administration, because the raw-score-to-scale-score conversion tables were developed before tests were administered using pre-equating, individual student results were available for download prior to the printing of paper reports. This electronic reporting was made possible through the QTR module in the Test Management System.

In order to score and report STS 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:

- **General Reporting Specifications**—Provides the calculation rules for the information presented on CAASPP summary reports and defines the appropriate codes to use when a student does not take or complete a test or when a score will not be reported
- **Score Key and Score Conversion**—Defines file formats and information that is provided for scoring and the process of converting raw scores to scale scores
- **Form Planner Specifications**—Describes, in detail, the contents of files that contain keys required for scoring
- **Aggregation Rules**—Describes how and when a school's results are aggregated at the school, district, county, and state levels
- **"What If" List**—Provides a variety of anomalous scenarios that may occur when test materials are returned by LEAs to Pearson and defines the action(s) to be taken in response
- **Edit Specifications**—Describes edits, defaults, and solutions to errors encountered while data are being captured as answer documents are processed
- **Reporting Cluster Names and Item Numbers**—Identifies the reporting clusters for each test and the number of items in each cluster

The scoring specifications are reviewed and revised by the CDE, ETS, and Pearson 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 Pearson 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. Pearson 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 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 answered correctly.

### Subscore

The items in each STS 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 reported both as raw scores and percent of items answered correctly. A description of the STS reporting clusters is provided in Appendix 2.B of Chapter 2, starting on page 19.

### Scale Score

Raw scores obtained on each STS are transformed to three-digit scale scores using the equating process described in Chapter 2 on page 13. Scale scores range from 150 to 600 on each STS. 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 is categorized into one of the following performance levels:

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

For all STS, 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, 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 Pearson. ETS independently generates scores and verifies Pearson's 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 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 scores.

### Individual Scores

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

#### Score Distributions and Summary Statistics

Summary statistics that describe the performance of students on each STS in the overall population are presented in Table 7.1. Analogous results are given in Table 7.2 and Table 7.3 for students in the target population and the optional population, respectively. The target population consists of students receiving instruction in Spanish or students who have attended school in the United States for less than 12 cumulative months. The 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 (see Chapter 1 on page 1 for more information on the intended population of STS test-takers). The optional population does *not* include students who were administered the STS for Non-ELs in Dual-Immersion Programs. In 2014, all STS test-takers in both the target and optional populations were considered voluntary.

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 (Overall Population)**

Content Area	STS *	No. of Items	No. of Examinees	Scale Score		Raw Score		Raw Score Percent Correct	
				Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Reading/ Language Arts	2	65	2,753	318	54	37.97	13.48	58.42	20.75
	3	65	1,878	311	44	32.70	11.14	50.30	17.13
	4	75	1,131	315	52	40.81	14.33	54.41	19.11
	5	75	848	309	58	35.77	13.50	47.69	18.00
	6	75	517	318	61	37.36	13.74	49.81	18.32
	7	75	496	322	57	39.44	13.44	52.58	17.92
	8	75	477	322	57	38.65	13.59	51.53	18.11
	9	75	1,027	319	54	40.73	12.86	54.31	17.15
	10	75	407	314	58	39.92	13.17	53.23	17.56
	11	75	198	319	61	40.92	12.57	54.57	16.77

\* Numbers indicate grade-level tests.

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

**Table 7.2 Mean and Standard Deviation of Raw and Scale Scores for the STS (Target Population)**

Content Area	STS *	No. of Items	No. of Examinees	Scale Score		Raw Score		Raw Score Percent Correct	
				Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Reading/ Language Arts	2	65	1,930	319	55	38.31	13.49	58.94	20.76
	3	65	1,388	312	45	32.78	11.21	50.42	17.25
	4	75	803	316	52	41.17	14.34	54.89	19.12
	5	75	645	313	59	36.80	13.75	49.07	18.34
	6	75	482	318	62	37.46	13.81	49.94	18.41
	7	75	460	323	57	39.59	13.54	52.79	18.05
	8	75	439	323	57	38.77	13.56	51.69	18.08
	9	75	956	320	54	40.79	12.89	54.38	17.18
	10	75	388	314	58	40.01	13.02	53.35	17.36
	11	75	183	319	61	41.02	12.57	54.70	16.76

\* Numbers indicate grade-level tests.

**Note:** 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.

**Table 7.3 Mean and Standard Deviation of Raw and Scale Scores for the STS (Optional Population)**

Content Area	STS *	No. of Items	No. of Examinees	Scale Score		Raw Score		Raw Score Percent Correct	
				Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Reading/ Language Arts	2	65	823	314	53	37.17	13.44	57.18	20.68
	3	65	490	310	43	32.47	10.93	49.96	16.81
	4	75	328	312	52	39.93	14.30	53.24	19.07
	5	75	203	295	52	32.46	12.12	43.28	16.16
	6	75	35	311	57	35.97	12.86	47.96	17.14
	7	75	36	313	50	37.44	12.19	49.93	16.26
	8	75	38	317	59	37.29	14.03	49.72	18.71
	9	75	71	315	51	39.96	12.54	53.28	16.72
	10	75	19	306	74	38.05	16.19	50.74	21.59
	11	75	15	314	67	39.73	13.00	52.98	17.34

\* Numbers indicate grade-level tests.

Statistics are not reported for tests with N less than 11.

**Note:** The 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.

The percentages of students in each performance level are presented in Table 7.4. 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 P1 data file was used for the analyses in this chapter. This file contained data collected from all LEAs but did not include corrections of demographic data through the California Longitudinal Pupil Achievement Data System (CALPADS). In addition, students with invalid scores were excluded from the tables.

**Table 7.4 Percentages of Examinees in Each Performance Level**

Content Area	STS *	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Proficient/ Advanced †
Reading/ Language Arts	2	8%	30%	33%	20%	10%	30%
	3	7%	36%	38%	15%	5%	20%
	4	15%	24%	32%	21%	8%	29%
	5	28%	19%	25%	19%	9%	28%
	6	20%	22%	26%	22%	10%	32%
	7	12%	25%	33%	21%	9%	30%
	8	9%	27%	33%	22%	9%	31%
	9	9%	27%	33%	22%	9%	31%
	10	12%	27%	32%	20%	9%	29%
	11	11%	21%	30%	28%	10%	38%

\* Numbers indicate grade-level tests.

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

Table 7.A.1 and Table 7.A.2 in Appendix 7.A, starting on page 112, show the distributions of scale scores for each STS for the overall and target populations, respectively.

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.

The distributions are presented for all examinees and for the target population 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 113 in Table 7.B.1 through Table 7.B.20 for the STS. The summary tables are provided for each STS based on all examinees and the target population, respectively. 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, country of origin, economic status, length of enrollment in U.S. schools, EL program participation, 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 7.5 provides definitions of the demographic groups included in the tables. Students' economic status was determined by considering the education level of their parents and whether or not they participated in the National School Lunch Program (NSLP).

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 7.5 Subgroup Definitions**

<b>Subgroup</b>	<b>Definition</b>
Gender	<ul style="list-style-type: none"> <li>• Male</li> <li>• Female</li> </ul>
Country of Origin	<ul style="list-style-type: none"> <li>• Argentina</li> <li>• Bolivia</li> <li>• Brazil</li> <li>• Chile</li> <li>• Colombia</li> <li>• Costa Rica</li> <li>• Cuba</li> <li>• Ecuador</li> <li>• El Salvador</li> <li>• Guatemala</li> <li>• Mexico</li> <li>• Nicaragua</li> <li>• Panama</li> <li>• Paraguay</li> <li>• Peru</li> <li>• Puerto Rico</li> <li>• Spain</li> <li>• United States</li> <li>• Uruguay</li> <li>• Venezuela</li> <li>• Other</li> </ul>

Subgroup	Definition
Economic Status	<ul style="list-style-type: none"> <li>• Not economically disadvantaged</li> <li>• Economically disadvantaged</li> </ul>
Enrollment in U.S. Schools	<ul style="list-style-type: none"> <li>• Less than 12 months</li> <li>• 12 months or more</li> </ul>
EL Program Participation	<ul style="list-style-type: none"> <li>• Primary Language Instruction and English-Language Development (ELD) Instruction and/or Specially Designed Academic Instruction in English (SDAIE) Instruction</li> <li>• ELD Instruction Only</li> <li>• SDAIE Instruction Only</li> <li>• ELD Instruction and SDAIE Instruction but Not Primary Language Instruction</li> <li>• Other EL Instructional Services</li> <li>• None (EL only)</li> </ul>
Special Services	<ul style="list-style-type: none"> <li>• No special services</li> <li>• Special services</li> </ul>

## Reports Produced and Scores for Each Report

The tests that make up the 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 reports. These categories and the specific reports in each category are given in Table 7.6.

**Table 7.6 Types of STS Reports**

1. Summary Reports	<ul style="list-style-type: none"> <li>▪ CAASPP Student Master List Summary</li> <li>▪ CAASPP Subgroup Summary</li> </ul>
2. Individual Reports	<ul style="list-style-type: none"> <li>▪ CAASPP Student Record Label</li> <li>▪ CAASPP Student Master List</li> <li>▪ CAASPP Student Report for the STS</li> </ul>
3. Internet Reports	<ul style="list-style-type: none"> <li>▪ STS Scores (state, county, LEA, school)</li> <li>▪ STS Summary Scores (state, county, LEA, school)</li> </ul>

These reports are sent to the independently testing charter schools, counties, or school districts; the LEA forwards the appropriate reports to test sites or, in the case of the CAASPP Student 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 Report, Student Record Label, and Student Master List 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 summary reports, student record labels, and the master lists via electronic reporting, accessed using the QTR module to the Test Management System. This module permits LEAs to download a file containing student data that includes scale scores and performance levels for all tests taken.

### **Score Report Contents**

The CAASPP Student Report provides scale scores, performance levels, and reporting cluster (subscore) 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. In addition, percent-correct scores are provided at the cluster level.

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 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 Report and the other reports is provided in Appendix 7.C on page 134.

### **Score Report Applications**

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

STS results help LEAs and schools identify strengths and weaknesses in their instructional programs. Each year, LEAs and school staffs examine STS 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 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 results (CDE, 2014a). 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, 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 *2014 CAASPP Post-Test Guide* (CDE, 2014b).

## References

California Department of Education. (2014a). *2014 CAASPP CST/CMA, CAPA, and STS printed reports*. Sacramento, CA. Retrieved from <http://caaspp.org/rsc/pdfs/caaspp.reports.2014.pdf>

California Department of Education. (2014b). *CAASPP 2014 post-test guide*. Sacramento, CA. Retrieved from [http://caaspp.org/rsc/pdfs/caaspp.post-test\\_guide.2014.pdf](http://caaspp.org/rsc/pdfs/caaspp.post-test_guide.2014.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.

In 2014, administration of the STS 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

**Table 7.A.1 Distribution of STS Scale Scores for RLA, 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	5	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	1	1	0	0	0	0	0
480 – 509	4	0	1	2	1	1	1	1	1	0
450 – 479	11	2	4	7	9	8	10	8	1	0
420 – 449	80	28	15	29	20	20	15	16	6	7
390 – 419	165	53	59	43	33	32	40	78	28	15
360 – 389	329	199	157	97	75	69	61	135	65	36
330 – 359	554	366	201	104	75	85	72	210	77	30
300 – 329	554	435	243	141	88	97	104	216	68	44
270 – 299	492	441	196	195	88	83	77	147	56	22
240 – 269	394	306	167	148	80	78	71	165	55	18
210 – 239	147	47	81	70	41	21	23	30	44	18
180 – 209	17	1	5	9	6	2	2	19	5	7
150 – 179	1	0	2	2	0	0	1	2	1	1

**Table 7.A.2 Distribution of STS Scale Scores for RLA, Grades Two through Eleven (Target 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	4	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	1	1	0	0	0	0	0
480 – 509	4	0	0	1	1	1	1	1	1	0
450 – 479	10	2	3	6	9	8	9	8	1	0
420 – 449	60	20	12	26	20	19	15	16	5	6
390 – 419	121	45	43	39	30	31	35	76	26	13
360 – 389	234	153	113	80	70	64	58	121	63	36
330 – 359	382	255	150	85	71	76	68	193	74	30
300 – 329	392	318	167	106	77	91	96	200	66	39
270 – 299	357	338	139	139	84	78	68	140	53	16
240 – 269	252	225	117	104	77	71	66	156	53	18
210 – 239	98	32	53	50	37	19	20	26	42	18
180 – 209	15	0	4	6	5	2	2	17	3	7
150 – 179	1	0	2	2	0	0	1	2	1	0

## 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, administration of the STS 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

**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					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
All population valid scores	2,753	318	54	8%	30%	33%	20%	10%	65%	57%	58%	58%	45%
Male	1,390	312	53	8%	35%	31%	17%	8%	63%	54%	55%	55%	43%
Female	1,355	324	55	7%	26%	34%	22%	11%	67%	60%	61%	60%	46%
Gender unknown	8	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	0	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	0	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	1	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	3	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	0	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	1	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	0	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	69	287	55	23%	43%	19%	9%	6%	52%	47%	47%	42%	36%
Guatemala	39	290	47	18%	38%	33%	10%	0%	54%	46%	48%	47%	36%
Mexico	246	308	58	11%	34%	30%	19%	6%	61%	55%	54%	52%	42%
Nicaragua	3	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	4	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	1	—	—	—	—	—	—	—	—	—	—	—	—
Spain	2	—	—	—	—	—	—	—	—	—	—	—	—
United States	2,228	320	54	7%	30%	33%	20%	10%	66%	57%	59%	59%	45%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	0	—	—	—	—	—	—	—	—	—	—	—	—
Other	22	288	55	23%	45%	23%	0%	9%	51%	44%	48%	44%	40%
Country unknown	134	329	54	5%	26%	31%	24%	14%	69%	61%	64%	62%	51%

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Not economically disadvantaged	111	311	57	10%	33%	30%	18%	9%	63%	53%	57%	55%	42%
Economically disadvantaged	2,515	318	54	8%	30%	33%	20%	10%	65%	57%	58%	58%	45%
Economic status unknown	127	321	51	5%	31%	34%	20%	10%	66%	58%	62%	59%	49%
In U.S. schools < 12 months	526	297	60	18%	40%	23%	10%	9%	56%	49%	50%	47%	39%
In U.S. schools ≥ 12 months	2,227	323	52	5%	28%	35%	22%	10%	67%	59%	60%	60%	46%
Primary language instruction and ELD and/or SDAIE instruction	1,642	327	53	4%	26%	35%	22%	12%	69%	60%	61%	62%	48%
ELD instruction only	20	271	35	25%	55%	15%	5%	0%	48%	38%	38%	35%	28%
SDAIE instruction only	211	267	38	31%	50%	16%	2%	1%	44%	37%	40%	33%	27%
ELD instr. and SDAIE instr. but not primary language instr.	747	312	52	8%	34%	32%	19%	7%	63%	54%	56%	55%	43%
Other EL instructional services	47	335	52	0%	26%	34%	30%	11%	69%	65%	70%	66%	49%
None (EL only)	60	327	51	5%	22%	40%	17%	17%	71%	60%	63%	61%	48%
Program participation unknown	26	288	54	27%	31%	31%	8%	4%	53%	45%	49%	43%	40%
No special education	2,644	319	54	8%	29%	33%	20%	10%	66%	57%	58%	58%	45%
Special education	109	287	44	13%	57%	21%	5%	5%	52%	45%	48%	44%	33%
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

**Table 7.B.2 Demographic Summary for RLA, Grade Two (Target Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target population valid scores	1,930	319	55	8%	30%	33%	20%	10%	66%	57%	59%	58%	45%
Male	979	313	53	8%	34%	31%	18%	8%	64%	54%	56%	56%	44%
Female	947	325	56	7%	25%	34%	21%	13%	68%	60%	61%	61%	47%
Gender unknown	4	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	0	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	0	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	0	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	3	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	0	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	1	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	0	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	57	284	52	21%	49%	19%	5%	5%	51%	44%	45%	41%	35%
Guatemala	32	292	44	19%	34%	41%	6%	0%	55%	47%	49%	48%	36%
Mexico	214	308	58	12%	34%	31%	18%	5%	60%	55%	53%	52%	42%
Nicaragua	3	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	4	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	1	—	—	—	—	—	—	—	—	—	—	—	—
Spain	2	—	—	—	—	—	—	—	—	—	—	—	—
United States	1,471	322	54	6%	29%	33%	20%	11%	67%	58%	60%	60%	46%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	0	—	—	—	—	—	—	—	—	—	—	—	—
Other	18	278	51	28%	50%	17%	0%	6%	47%	41%	45%	38%	37%
Country unknown	124	330	55	6%	24%	31%	25%	15%	69%	61%	64%	62%	52%
Not economically disadvantaged	87	309	58	13%	32%	29%	17%	9%	62%	52%	56%	54%	42%
Economically disadvantaged	1,720	320	55	8%	30%	33%	20%	10%	66%	57%	58%	58%	45%
Economic status unknown	123	322	51	4%	32%	34%	20%	11%	66%	58%	62%	59%	50%
In U.S. schools < 12 months	526	297	60	18%	40%	23%	10%	9%	56%	49%	50%	47%	39%
In U.S. schools ≥ 12 months	1,404	327	51	4%	26%	36%	23%	11%	69%	61%	62%	63%	48%
Primary language instruction and ELD and/or SDAIE instruction	1,642	327	53	4%	26%	35%	22%	12%	69%	60%	61%	62%	48%
ELD instruction only	12	271	38	33%	42%	17%	8%	0%	48%	42%	35%	32%	29%
SDAIE instruction only	140	270	38	27%	53%	17%	1%	1%	45%	38%	42%	34%	29%
ELD instr. and SDAIE instr. but not primary language instr.	117	281	47	21%	48%	21%	9%	2%	51%	43%	47%	37%	34%
Other EL instructional services	2	—	—	—	—	—	—	—	—	—	—	—	—
None (EL only)	8	—	—	—	—	—	—	—	—	—	—	—	—
Program participation unknown	9	—	—	—	—	—	—	—	—	—	—	—	—
No special education	1,852	321	55	7%	29%	33%	20%	11%	66%	58%	59%	59%	46%
Special education	78	284	42	10%	62%	22%	4%	3%	51%	44%	49%	41%	31%
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

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

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
All population valid scores	1,878	311	44	8%	35%	39%	14%	4%	57%	47%	49%	48%	46%
Male	944	305	44	9%	40%	35%	11%	4%	55%	44%	48%	45%	44%
Female	932	317	44	6%	29%	42%	17%	5%	59%	49%	51%	51%	48%
Gender unknown	2	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	1	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	0	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	0	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	1	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	0	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	1	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	0	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	109	309	46	7%	40%	35%	14%	4%	59%	46%	53%	42%	42%
Guatemala	60	309	42	5%	42%	37%	10%	7%	59%	44%	49%	45%	44%
Mexico	197	312	48	8%	36%	35%	15%	6%	57%	47%	51%	46%	47%
Nicaragua	2	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	3	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	0	—	—	—	—	—	—	—	—	—	—	—	—
Spain	1	—	—	—	—	—	—	—	—	—	—	—	—
United States	1,364	311	44	8%	34%	40%	14%	4%	56%	47%	49%	49%	46%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	0	—	—	—	—	—	—	—	—	—	—	—	—
Other	22	293	47	14%	55%	18%	9%	5%	48%	40%	51%	38%	36%
Country unknown	117	320	38	4%	27%	48%	17%	3%	62%	49%	52%	50%	51%
Not economically disadvantaged	91	334	49	3%	22%	42%	18%	15%	66%	55%	61%	56%	53%
Economically disadvantaged	1,706	310	44	8%	36%	38%	14%	4%	56%	46%	49%	47%	45%
Economic status unknown	81	315	37	5%	30%	51%	12%	2%	60%	47%	50%	49%	49%
In U.S. schools < 12 months	466	306	46	9%	39%	35%	13%	5%	56%	45%	50%	43%	43%
In U.S. schools ≥ 12 months	1,412	313	44	7%	33%	40%	15%	4%	57%	47%	49%	49%	47%
Primary language instruction and ELD and/or SDAIE instruction	1,117	315	44	6%	33%	39%	16%	5%	59%	48%	50%	50%	48%
ELD instruction only	29	305	48	14%	31%	41%	7%	7%	52%	42%	53%	45%	47%
SDAIE instruction only	167	290	42	17%	49%	23%	10%	2%	49%	39%	44%	37%	37%
ELD instr. and SDAIE instr. but not primary language instr.	477	308	42	8%	35%	43%	11%	3%	55%	46%	49%	46%	45%
Other EL instructional services	28	333	48	4%	25%	39%	25%	7%	66%	54%	56%	55%	56%
None (EL only)	42	321	41	2%	31%	40%	21%	5%	64%	47%	56%	49%	50%
Program participation unknown	18	298	46	22%	22%	44%	11%	0%	53%	43%	52%	41%	34%
No special education	1,815	312	44	7%	34%	40%	15%	5%	57%	47%	50%	48%	46%
Special education	63	278	36	27%	52%	19%	2%	0%	42%	35%	38%	36%	35%
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

**Table 7.B.4 Demographic Summary for RLA, Grade Three (Target Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target population valid scores	1,388	312	45	7%	36%	38%	15%	5%	57%	47%	49%	48%	46%
Male	695	306	44	7%	41%	35%	12%	5%	56%	44%	48%	45%	44%
Female	691	317	45	7%	30%	40%	18%	5%	59%	49%	51%	51%	48%
Gender unknown	2	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	1	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	0	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	0	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	1	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	0	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	1	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	0	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	84	310	44	6%	40%	37%	13%	4%	59%	47%	54%	41%	41%
Guatemala	51	309	40	6%	39%	39%	12%	4%	59%	45%	50%	43%	44%
Mexico	162	309	48	8%	40%	34%	13%	6%	57%	46%	50%	44%	45%
Nicaragua	2	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	3	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	0	—	—	—	—	—	—	—	—	—	—	—	—
Spain	1	—	—	—	—	—	—	—	—	—	—	—	—
United States	954	312	45	7%	35%	37%	15%	5%	57%	47%	48%	49%	46%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	0	—	—	—	—	—	—	—	—	—	—	—	—
Other	18	286	36	11%	61%	17%	11%	0%	45%	38%	47%	36%	35%
Country unknown	110	321	38	4%	27%	47%	18%	4%	62%	50%	53%	50%	52%
Not economically disadvantaged	72	333	50	4%	21%	44%	15%	15%	66%	54%	60%	55%	53%
Economically disadvantaged	1,242	310	45	8%	37%	36%	15%	4%	57%	46%	49%	47%	46%
Economic status unknown	74	316	36	3%	32%	50%	12%	3%	60%	47%	51%	50%	49%
In U.S. schools < 12 months	466	306	46	9%	39%	35%	13%	5%	56%	45%	50%	43%	43%
In U.S. schools ≥ 12 months	922	314	44	6%	34%	39%	16%	5%	58%	47%	49%	50%	48%
Primary language instruction and ELD and/or SDAIE instruction	1,117	315	44	6%	33%	39%	16%	5%	59%	48%	50%	50%	48%
ELD instruction only	11	293	48	9%	55%	27%	0%	9%	50%	38%	50%	38%	37%
SDAIE instruction only	117	294	40	11%	51%	25%	13%	0%	51%	41%	46%	37%	39%
ELD instr. and SDAIE instr. but not primary language instr.	123	298	44	12%	42%	35%	7%	4%	52%	41%	48%	39%	42%
Other EL instructional services	1	—	—	—	—	—	—	—	—	—	—	—	—
None (EL only)	13	309	38	0%	46%	31%	23%	0%	57%	43%	56%	42%	49%
Program participation unknown	6	—	—	—	—	—	—	—	—	—	—	—	—
No special education	1,337	313	45	7%	35%	38%	15%	5%	58%	47%	50%	48%	47%
Special education	51	278	35	24%	57%	18%	2%	0%	43%	35%	39%	35%	34%
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

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

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
All population valid scores	1,131	315	52	16%	24%	32%	19%	8%	61%	55%	48%	58%	46%
Male	578	304	52	20%	30%	29%	13%	8%	57%	51%	45%	54%	42%
Female	551	325	50	11%	18%	36%	26%	9%	65%	58%	53%	62%	50%
Gender unknown	2	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	0	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	0	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	0	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	6	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	1	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	0	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	2	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	90	302	48	16%	27%	43%	12%	2%	58%	49%	46%	54%	41%
Guatemala	33	306	47	15%	30%	39%	12%	3%	60%	50%	45%	55%	43%
Mexico	203	314	54	17%	26%	28%	23%	7%	59%	57%	50%	57%	45%
Nicaragua	3	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	3	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	0	—	—	—	—	—	—	—	—	—	—	—	—
Spain	0	—	—	—	—	—	—	—	—	—	—	—	—
United States	715	317	52	15%	23%	32%	20%	10%	62%	55%	49%	59%	47%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	0	—	—	—	—	—	—	—	—	—	—	—	—
Other	26	312	47	19%	19%	38%	15%	8%	60%	59%	41%	56%	45%
Country unknown	49	313	53	16%	24%	33%	20%	6%	59%	54%	48%	56%	48%
Not economically disadvantaged	66	319	58	15%	20%	39%	14%	12%	62%	58%	53%	57%	47%
Economically disadvantaged	1,010	315	52	16%	24%	32%	20%	8%	61%	55%	48%	58%	46%
Economic status unknown	55	308	51	18%	25%	35%	16%	5%	56%	53%	47%	54%	46%
In U.S. schools < 12 months	470	310	51	17%	27%	32%	16%	7%	58%	54%	48%	56%	43%
In U.S. schools ≥ 12 months	661	318	52	15%	22%	33%	22%	9%	62%	55%	49%	59%	48%
Primary language instruction and ELD and/or SDAIE instruction	491	324	51	12%	21%	33%	24%	11%	65%	58%	52%	62%	49%
ELD instruction only	24	315	46	17%	21%	38%	25%	0%	63%	55%	50%	57%	46%
SDAIE instruction only	135	292	53	30%	31%	25%	7%	7%	51%	47%	43%	48%	38%
ELD instr. and SDAIE instr. but not primary language instr.	419	309	49	16%	28%	32%	18%	5%	58%	54%	46%	56%	45%
Other EL instructional services	2	—	—	—	—	—	—	—	—	—	—	—	—
None (EL only)	35	341	50	6%	9%	46%	11%	29%	71%	64%	57%	70%	54%
Program participation unknown	25	299	50	24%	16%	44%	12%	4%	56%	50%	43%	54%	38%
No special education	1,097	316	52	15%	24%	33%	20%	8%	61%	55%	49%	58%	46%
Special education	34	282	48	38%	29%	21%	9%	3%	48%	42%	34%	46%	35%
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

**Table 7.B.6 Demographic Summary for RLA, Grade Four (Target Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target population valid scores	803	316	52	15%	24%	32%	21%	8%	61%	55%	49%	58%	46%
Male	419	306	52	19%	32%	27%	14%	9%	58%	52%	45%	54%	42%
Female	383	326	49	11%	16%	37%	28%	8%	65%	59%	54%	63%	50%
Gender unknown	1	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	0	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	0	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	0	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	5	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	0	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	0	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	2	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	77	299	50	18%	30%	38%	12%	3%	56%	48%	45%	53%	40%
Guatemala	27	293	39	19%	37%	44%	0%	0%	56%	44%	41%	49%	38%
Mexico	172	314	51	16%	26%	27%	24%	6%	59%	58%	51%	56%	45%
Nicaragua	3	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	2	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	0	—	—	—	—	—	—	—	—	—	—	—	—
Spain	0	—	—	—	—	—	—	—	—	—	—	—	—
United States	483	321	53	14%	21%	31%	22%	11%	63%	56%	51%	61%	48%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	0	—	—	—	—	—	—	—	—	—	—	—	—
Other	24	307	45	21%	21%	42%	13%	4%	59%	58%	38%	54%	43%
Country unknown	8	—	—	—	—	—	—	—	—	—	—	—	—
Not economically disadvantaged	57	320	58	14%	19%	40%	14%	12%	63%	58%	54%	58%	47%
Economically disadvantaged	734	316	51	15%	24%	31%	21%	8%	61%	55%	49%	59%	46%
Economic status unknown	12	295	42	25%	33%	25%	17%	0%	50%	51%	46%	46%	42%
In U.S. schools < 12 months	470	310	51	17%	27%	32%	16%	7%	58%	54%	48%	56%	43%
In U.S. schools ≥ 12 months	333	324	52	13%	19%	31%	27%	10%	66%	57%	51%	62%	49%
Primary language instruction and ELD and/or SDAIE instruction	491	324	51	12%	21%	33%	24%	11%	65%	58%	52%	62%	49%
ELD instruction only	8	—	—	—	—	—	—	—	—	—	—	—	—
SDAIE instruction only	102	295	52	24%	34%	29%	6%	7%	52%	49%	44%	49%	39%
ELD instr. and SDAIE instr. but not primary language instr.	178	307	50	19%	28%	28%	21%	4%	57%	54%	47%	54%	43%
Other EL instructional services	1	—	—	—	—	—	—	—	—	—	—	—	—
None (EL only)	10	—	—	—	—	—	—	—	—	—	—	—	—
Program participation unknown	13	304	48	23%	15%	46%	8%	8%	59%	55%	47%	52%	36%
No special education	778	317	52	15%	24%	32%	21%	8%	62%	56%	50%	59%	46%
Special education	25	282	52	40%	28%	16%	12%	4%	48%	42%	36%	47%	34%
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

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

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
All population valid scores	848	309	58	30%	20%	25%	18%	7%	47%	44%	46%	54%	46%
Male	457	301	56	36%	19%	23%	16%	5%	45%	42%	44%	51%	44%
Female	387	318	59	23%	20%	27%	20%	10%	49%	47%	49%	57%	49%
Gender unknown	4	–	–	–	–	–	–	–	–	–	–	–	–
Argentina	0	–	–	–	–	–	–	–	–	–	–	–	–
Bolivia	0	–	–	–	–	–	–	–	–	–	–	–	–
Brazil	0	–	–	–	–	–	–	–	–	–	–	–	–
Chile	0	–	–	–	–	–	–	–	–	–	–	–	–
Colombia	2	–	–	–	–	–	–	–	–	–	–	–	–
Costa Rica	2	–	–	–	–	–	–	–	–	–	–	–	–
Cuba	2	–	–	–	–	–	–	–	–	–	–	–	–
Ecuador	0	–	–	–	–	–	–	–	–	–	–	–	–
El Salvador	90	303	54	33%	22%	21%	20%	3%	47%	42%	47%	51%	43%
Guatemala	44	303	55	32%	18%	30%	18%	2%	44%	42%	46%	54%	44%
Mexico	226	320	62	24%	18%	27%	19%	11%	50%	47%	53%	56%	50%
Nicaragua	2	–	–	–	–	–	–	–	–	–	–	–	–
Panama	0	–	–	–	–	–	–	–	–	–	–	–	–
Paraguay	0	–	–	–	–	–	–	–	–	–	–	–	–
Peru	2	–	–	–	–	–	–	–	–	–	–	–	–
Puerto Rico	0	–	–	–	–	–	–	–	–	–	–	–	–
Spain	1	–	–	–	–	–	–	–	–	–	–	–	–
United States	434	303	56	34%	20%	22%	18%	6%	46%	43%	43%	53%	45%
Uruguay	0	–	–	–	–	–	–	–	–	–	–	–	–
Venezuela	0	–	–	–	–	–	–	–	–	–	–	–	–
Other	27	319	58	22%	15%	37%	15%	11%	48%	44%	52%	55%	54%
Country unknown	16	312	64	25%	19%	31%	13%	13%	48%	45%	46%	56%	48%
Not economically disadvantaged	53	321	55	19%	19%	28%	25%	9%	49%	47%	53%	56%	52%
Economically disadvantaged	760	308	59	31%	19%	24%	18%	7%	47%	44%	46%	54%	46%
Economic status unknown	35	300	41	26%	31%	29%	14%	0%	46%	37%	43%	51%	47%
In U.S. schools < 12 months	450	315	60	27%	18%	26%	20%	9%	49%	46%	51%	55%	48%
In U.S. schools ≥ 12 months	398	301	55	34%	21%	23%	16%	5%	45%	42%	41%	53%	44%
Primary language instruction and ELD and/or SDAIE instruction	294	314	57	26%	21%	26%	18%	9%	48%	47%	46%	57%	47%
ELD instruction only	35	297	58	34%	20%	31%	9%	6%	46%	38%	44%	50%	43%
SDAIE instruction only	151	301	57	34%	21%	23%	17%	5%	46%	42%	45%	50%	44%
ELD instr. and SDAIE instr. but not primary language instr.	306	307	57	33%	19%	25%	17%	6%	46%	43%	47%	52%	47%
Other EL instructional services	4	–	–	–	–	–	–	–	–	–	–	–	–
None (EL only)	13	312	59	31%	15%	23%	31%	0%	42%	48%	53%	52%	50%
Program participation unknown	45	315	67	31%	16%	18%	27%	9%	49%	45%	48%	57%	48%
No special education	818	310	58	30%	19%	25%	19%	7%	47%	45%	47%	54%	47%
Special education	30	276	51	50%	23%	20%	3%	3%	36%	33%	33%	46%	38%
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

**Table 7.B.8 Demographic Summary for RLA, Grade Five (Target Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target population valid scores	645	313	59	28%	19%	25%	19%	9%	48%	46%	49%	55%	47%
Male	348	304	57	34%	19%	23%	18%	6%	46%	43%	46%	51%	45%
Female	293	324	59	19%	19%	28%	22%	12%	51%	49%	52%	59%	51%
Gender unknown	4	–	–	–	–	–	–	–	–	–	–	–	–
Argentina	0	–	–	–	–	–	–	–	–	–	–	–	–
Bolivia	0	–	–	–	–	–	–	–	–	–	–	–	–
Brazil	0	–	–	–	–	–	–	–	–	–	–	–	–
Chile	0	–	–	–	–	–	–	–	–	–	–	–	–
Colombia	2	–	–	–	–	–	–	–	–	–	–	–	–
Costa Rica	1	–	–	–	–	–	–	–	–	–	–	–	–
Cuba	2	–	–	–	–	–	–	–	–	–	–	–	–
Ecuador	0	–	–	–	–	–	–	–	–	–	–	–	–
El Salvador	70	302	56	34%	20%	23%	20%	3%	46%	43%	47%	51%	42%
Guatemala	34	306	56	29%	21%	26%	21%	3%	45%	43%	47%	56%	44%
Mexico	187	321	63	23%	18%	28%	19%	12%	50%	47%	54%	56%	50%
Nicaragua	2	–	–	–	–	–	–	–	–	–	–	–	–
Panama	0	–	–	–	–	–	–	–	–	–	–	–	–
Paraguay	0	–	–	–	–	–	–	–	–	–	–	–	–
Peru	2	–	–	–	–	–	–	–	–	–	–	–	–
Puerto Rico	0	–	–	–	–	–	–	–	–	–	–	–	–
Spain	1	–	–	–	–	–	–	–	–	–	–	–	–
United States	312	310	57	30%	20%	22%	21%	8%	48%	46%	46%	54%	47%
Uruguay	0	–	–	–	–	–	–	–	–	–	–	–	–
Venezuela	0	–	–	–	–	–	–	–	–	–	–	–	–
Other	24	315	60	25%	17%	38%	8%	13%	47%	43%	52%	53%	53%
Country unknown	8	–	–	–	–	–	–	–	–	–	–	–	–
Not economically disadvantaged	45	322	55	16%	22%	29%	24%	9%	50%	48%	54%	57%	51%
Economically disadvantaged	595	312	59	29%	18%	25%	19%	9%	48%	46%	48%	54%	47%
Economic status unknown	5	–	–	–	–	–	–	–	–	–	–	–	–
In U.S. schools < 12 months	450	315	60	27%	18%	26%	20%	9%	49%	46%	51%	55%	48%
In U.S. schools ≥ 12 months	195	308	57	30%	21%	24%	17%	8%	46%	47%	43%	54%	45%
Primary language instruction and ELD and/or SDAIE instruction	294	314	57	26%	21%	26%	18%	9%	48%	47%	46%	57%	47%
ELD instruction only	18	300	66	39%	11%	33%	6%	11%	46%	38%	48%	47%	46%
SDAIE instruction only	113	307	60	33%	19%	23%	19%	7%	48%	43%	49%	52%	44%
ELD instr. and SDAIE instr. but not primary language instr.	191	318	61	26%	17%	27%	22%	9%	50%	47%	53%	54%	50%
Other EL instructional services	0	–	–	–	–	–	–	–	–	–	–	–	–
None (EL only)	12	309	60	33%	17%	17%	33%	0%	40%	48%	51%	50%	49%
Program participation unknown	17	301	59	35%	18%	24%	18%	6%	46%	39%	52%	47%	44%
No special education	627	314	59	27%	18%	26%	20%	9%	49%	46%	49%	55%	48%
Special education	18	277	40	44%	33%	17%	6%	0%	36%	34%	29%	47%	39%
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

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

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
All population valid scores	517	318	61	20%	21%	27%	21%	10%	46%	50%	56%	55%	44%
Male	265	308	60	26%	23%	27%	15%	9%	43%	48%	52%	51%	41%
Female	249	328	60	15%	20%	27%	27%	11%	48%	52%	59%	59%	47%
Gender unknown	3	–	–	–	–	–	–	–	–	–	–	–	–
Argentina	2	–	–	–	–	–	–	–	–	–	–	–	–
Bolivia	1	–	–	–	–	–	–	–	–	–	–	–	–
Brazil	1	–	–	–	–	–	–	–	–	–	–	–	–
Chile	0	–	–	–	–	–	–	–	–	–	–	–	–
Colombia	6	–	–	–	–	–	–	–	–	–	–	–	–
Costa Rica	0	–	–	–	–	–	–	–	–	–	–	–	–
Cuba	3	–	–	–	–	–	–	–	–	–	–	–	–
Ecuador	1	–	–	–	–	–	–	–	–	–	–	–	–
El Salvador	95	296	58	34%	20%	27%	15%	4%	40%	43%	52%	46%	38%
Guatemala	38	299	64	32%	32%	21%	8%	8%	38%	41%	54%	50%	38%
Mexico	158	329	63	17%	18%	27%	22%	16%	50%	54%	58%	57%	47%
Nicaragua	4	–	–	–	–	–	–	–	–	–	–	–	–
Panama	0	–	–	–	–	–	–	–	–	–	–	–	–
Paraguay	0	–	–	–	–	–	–	–	–	–	–	–	–
Peru	3	–	–	–	–	–	–	–	–	–	–	–	–
Puerto Rico	1	–	–	–	–	–	–	–	–	–	–	–	–
Spain	2	–	–	–	–	–	–	–	–	–	–	–	–
United States	174	324	58	14%	21%	28%	28%	9%	48%	52%	56%	59%	46%
Uruguay	0	–	–	–	–	–	–	–	–	–	–	–	–
Venezuela	0	–	–	–	–	–	–	–	–	–	–	–	–
Other	20	295	39	25%	40%	20%	15%	0%	38%	46%	50%	43%	37%
Country unknown	8	–	–	–	–	–	–	–	–	–	–	–	–
Not economically disadvantaged	42	309	65	29%	19%	33%	7%	12%	44%	47%	54%	51%	40%
Economically disadvantaged	461	319	61	20%	21%	26%	23%	10%	46%	50%	56%	55%	44%
Economic status unknown	14	295	48	21%	36%	29%	7%	7%	38%	49%	43%	45%	37%
In U.S. schools < 12 months	365	315	64	24%	22%	24%	19%	12%	45%	49%	56%	53%	43%
In U.S. schools ≥ 12 months	152	324	54	13%	20%	32%	28%	7%	48%	52%	55%	59%	46%
Primary language instruction and ELD and/or SDAIE instruction	161	329	56	10%	25%	27%	29%	9%	49%	54%	56%	60%	46%
ELD instruction only	8	–	–	–	–	–	–	–	–	–	–	–	–
SDAIE instruction only	24	323	56	13%	21%	42%	17%	8%	41%	57%	61%	54%	44%
ELD instr. and SDAIE instr. but not primary language instr.	302	314	64	25%	19%	25%	19%	12%	45%	48%	56%	53%	43%
Other EL instructional services	3	–	–	–	–	–	–	–	–	–	–	–	–
None (EL only)	11	277	57	55%	27%	9%	0%	9%	34%	37%	45%	38%	34%
Program participation unknown	8	–	–	–	–	–	–	–	–	–	–	–	–
No special education	508	319	61	20%	21%	27%	22%	10%	46%	50%	56%	55%	44%
Special education	9	–	–	–	–	–	–	–	–	–	–	–	–
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

**Table 7.B.10 Demographic Summary for RLA, Grade Six (Target Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target population valid scores	482	318	62	20%	22%	26%	22%	10%	46%	50%	56%	55%	44%
Male	243	307	61	26%	23%	26%	15%	10%	43%	48%	52%	50%	41%
Female	236	329	60	14%	21%	26%	28%	11%	49%	52%	60%	59%	47%
Gender unknown	3	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	2	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	1	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	1	—	—	—	—	—	—	—	—	—	—	—	—
Chile	0	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	6	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	0	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	3	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	1	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	92	296	59	34%	20%	27%	15%	4%	40%	43%	52%	46%	38%
Guatemala	36	301	65	31%	31%	22%	8%	8%	39%	41%	55%	51%	38%
Mexico	148	328	64	18%	20%	26%	22%	16%	49%	54%	58%	57%	47%
Nicaragua	4	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	2	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	1	—	—	—	—	—	—	—	—	—	—	—	—
Spain	1	—	—	—	—	—	—	—	—	—	—	—	—
United States	158	327	57	12%	22%	28%	29%	9%	48%	53%	57%	59%	47%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	0	—	—	—	—	—	—	—	—	—	—	—	—
Other	19	293	40	26%	42%	16%	16%	0%	39%	47%	50%	41%	36%
Country unknown	7	—	—	—	—	—	—	—	—	—	—	—	—
Not economically disadvantaged	41	308	65	29%	20%	32%	7%	12%	44%	47%	53%	50%	40%
Economically disadvantaged	429	320	62	19%	22%	25%	24%	10%	46%	51%	57%	56%	45%
Economic status unknown	12	284	35	25%	42%	25%	8%	0%	37%	46%	39%	43%	31%
In U.S. schools < 12 months	365	315	64	24%	22%	24%	19%	12%	45%	49%	56%	53%	43%
In U.S. schools ≥ 12 months	117	328	53	9%	23%	30%	31%	7%	49%	53%	55%	61%	47%
Primary language instruction and ELD and/or SDAIE instruction	161	329	56	10%	25%	27%	29%	9%	49%	54%	56%	60%	46%
ELD instruction only	5	—	—	—	—	—	—	—	—	—	—	—	—
SDAIE instruction only	19	335	54	5%	21%	42%	21%	11%	43%	61%	64%	59%	49%
ELD instr. and SDAIE instr. but not primary language instr.	282	314	65	26%	19%	25%	18%	12%	45%	48%	56%	53%	43%
Other EL instructional services	1	—	—	—	—	—	—	—	—	—	—	—	—
None (EL only)	11	277	57	55%	27%	9%	0%	9%	34%	37%	45%	38%	34%
Program participation unknown	3	—	—	—	—	—	—	—	—	—	—	—	—
No special education	474	319	62	20%	22%	26%	22%	11%	46%	50%	56%	55%	44%
Special education	8	—	—	—	—	—	—	—	—	—	—	—	—
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

**Table 7.B.11 Demographic Summary for RLA, Grade Seven (Overall Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
All population valid scores	496	322	57	11%	26%	33%	21%	9%	56%	49%	56%	53%	51%
Male	242	318	57	14%	26%	32%	21%	7%	56%	49%	55%	50%	50%
Female	247	326	57	9%	25%	34%	21%	11%	57%	50%	57%	55%	53%
Gender unknown	7	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	0	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	1	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	1	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	4	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	1	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	1	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	1	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	91	313	56	11%	36%	25%	21%	7%	54%	47%	53%	47%	50%
Guatemala	51	313	58	18%	29%	27%	18%	8%	53%	46%	51%	51%	49%
Mexico	156	331	56	8%	21%	38%	24%	10%	60%	52%	59%	56%	53%
Nicaragua	2	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	0	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	0	—	—	—	—	—	—	—	—	—	—	—	—
Spain	0	—	—	—	—	—	—	—	—	—	—	—	—
United States	151	325	58	15%	18%	35%	23%	10%	57%	49%	57%	55%	52%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	0	—	—	—	—	—	—	—	—	—	—	—	—
Other	26	292	42	15%	54%	19%	8%	4%	45%	43%	45%	42%	42%
Country unknown	10	—	—	—	—	—	—	—	—	—	—	—	—
Not economically disadvantaged	40	318	66	15%	30%	28%	20%	8%	55%	49%	54%	50%	50%
Economically disadvantaged	441	322	56	11%	25%	34%	21%	9%	57%	49%	56%	53%	51%
Economic status unknown	15	322	50	13%	20%	33%	27%	7%	48%	50%	56%	55%	53%
In U.S. schools < 12 months	390	326	58	9%	27%	31%	22%	11%	58%	51%	57%	53%	53%
In U.S. schools ≥ 12 months	106	306	48	20%	19%	42%	18%	2%	49%	44%	52%	51%	46%
Primary language instruction and ELD and/or SDAIE instruction	137	319	55	18%	17%	40%	18%	7%	54%	49%	54%	54%	50%
ELD instruction only	11	292	55	27%	27%	36%	9%	0%	49%	39%	44%	46%	42%
SDAIE instruction only	8	—	—	—	—	—	—	—	—	—	—	—	—
ELD instr. and SDAIE instr. but not primary language instr.	311	324	58	9%	29%	30%	22%	10%	58%	50%	57%	53%	52%
Other EL instructional services	3	—	—	—	—	—	—	—	—	—	—	—	—
None (EL only)	9	—	—	—	—	—	—	—	—	—	—	—	—
Program participation unknown	17	311	51	12%	29%	35%	24%	0%	56%	48%	49%	50%	46%
No special education	490	323	57	11%	26%	34%	21%	9%	57%	50%	56%	53%	51%
Special education	6	—	—	—	—	—	—	—	—	—	—	—	—
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

**Table 7.B.12 Demographic Summary for RLA, Grade Seven (Target Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Scale Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target population valid scores	460	323	57	12%	25%	33%	21%	9%	57%	50%	56%	53%	51%
Male	229	319	57	14%	26%	32%	21%	7%	56%	49%	55%	50%	50%
Female	225	327	58	9%	24%	34%	21%	11%	58%	50%	56%	55%	54%
Gender unknown	6	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	0	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	1	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	1	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	4	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	0	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	1	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	1	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	86	315	55	9%	36%	27%	21%	7%	55%	48%	54%	48%	50%
Guatemala	50	313	59	18%	30%	26%	18%	8%	53%	46%	51%	51%	49%
Mexico	143	333	57	8%	20%	38%	24%	10%	61%	53%	59%	57%	54%
Nicaragua	0	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	0	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	0	—	—	—	—	—	—	—	—	—	—	—	—
Spain	0	—	—	—	—	—	—	—	—	—	—	—	—
United States	140	324	59	15%	18%	36%	21%	10%	57%	49%	57%	55%	52%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	0	—	—	—	—	—	—	—	—	—	—	—	—
Other	25	293	42	16%	52%	20%	8%	4%	46%	43%	45%	42%	42%
Country unknown	8	—	—	—	—	—	—	—	—	—	—	—	—
Not economically disadvantaged	36	319	67	17%	25%	31%	19%	8%	56%	50%	54%	50%	49%
Economically disadvantaged	413	323	56	11%	26%	34%	21%	9%	57%	50%	56%	53%	52%
Economic status unknown	11	328	56	18%	18%	18%	36%	9%	50%	53%	56%	59%	56%
In U.S. schools < 12 months	390	326	58	9%	27%	31%	22%	11%	58%	51%	57%	53%	53%
In U.S. schools ≥ 12 months	70	303	47	24%	14%	44%	16%	1%	48%	42%	50%	51%	45%
Primary language instruction and ELD and/or SDAIE instruction	137	319	55	18%	17%	40%	18%	7%	54%	49%	54%	54%	50%
ELD instruction only	11	292	55	27%	27%	36%	9%	0%	49%	39%	44%	46%	42%
SDAIE instruction only	3	—	—	—	—	—	—	—	—	—	—	—	—
ELD instr. and SDAIE instr. but not primary language instr.	287	325	58	8%	29%	30%	23%	10%	59%	50%	57%	53%	53%
Other EL instructional services	2	—	—	—	—	—	—	—	—	—	—	—	—
None (EL only)	9	—	—	—	—	—	—	—	—	—	—	—	—
Program participation unknown	11	312	43	9%	27%	45%	18%	0%	57%	52%	45%	51%	44%
No special education	455	323	57	11%	25%	33%	21%	9%	57%	50%	56%	53%	52%
Special education	5	—	—	—	—	—	—	—	—	—	—	—	—
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

**Table 7.B.13 Demographic Summary for RLA, Grade Eight (Overall Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Number Correct Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
All population valid scores	477	322	57	9%	27%	32%	22%	9%	54%	47%	51%	61%	46%
Male	261	316	57	12%	28%	32%	20%	8%	54%	46%	50%	58%	43%
Female	213	331	55	4%	27%	33%	25%	11%	56%	48%	53%	66%	50%
Gender unknown	3	–	–	–	–	–	–	–	–	–	–	–	–
Argentina	0	–	–	–	–	–	–	–	–	–	–	–	–
Bolivia	0	–	–	–	–	–	–	–	–	–	–	–	–
Brazil	0	–	–	–	–	–	–	–	–	–	–	–	–
Chile	0	–	–	–	–	–	–	–	–	–	–	–	–
Colombia	5	–	–	–	–	–	–	–	–	–	–	–	–
Costa Rica	0	–	–	–	–	–	–	–	–	–	–	–	–
Cuba	1	–	–	–	–	–	–	–	–	–	–	–	–
Ecuador	0	–	–	–	–	–	–	–	–	–	–	–	–
El Salvador	112	307	52	14%	29%	38%	13%	5%	47%	43%	46%	57%	40%
Guatemala	41	294	50	20%	41%	24%	12%	2%	38%	41%	42%	49%	39%
Mexico	168	339	55	3%	24%	31%	28%	14%	62%	51%	55%	68%	51%
Nicaragua	3	–	–	–	–	–	–	–	–	–	–	–	–
Panama	0	–	–	–	–	–	–	–	–	–	–	–	–
Paraguay	0	–	–	–	–	–	–	–	–	–	–	–	–
Peru	7	–	–	–	–	–	–	–	–	–	–	–	–
Puerto Rico	1	–	–	–	–	–	–	–	–	–	–	–	–
Spain	1	–	–	–	–	–	–	–	–	–	–	–	–
United States	104	330	58	10%	19%	35%	26%	11%	58%	48%	54%	63%	50%
Uruguay	0	–	–	–	–	–	–	–	–	–	–	–	–
Venezuela	2	–	–	–	–	–	–	–	–	–	–	–	–
Other	23	286	46	13%	43%	39%	4%	0%	36%	39%	36%	48%	40%
Country unknown	9	–	–	–	–	–	–	–	–	–	–	–	–
Not economically disadvantaged	47	329	52	11%	19%	32%	30%	9%	58%	47%	57%	62%	49%
Economically disadvantaged	417	321	57	9%	29%	32%	21%	9%	54%	47%	50%	61%	46%
Economic status unknown	13	333	50	0%	23%	54%	15%	8%	50%	47%	52%	69%	53%
In U.S. schools < 12 months	416	324	57	8%	27%	33%	22%	10%	55%	47%	51%	62%	47%
In U.S. schools ≥ 12 months	61	309	54	13%	33%	30%	20%	5%	51%	43%	47%	56%	41%
Primary language instruction and ELD and/or SDAIE instruction	81	324	54	10%	20%	40%	23%	7%	58%	42%	52%	64%	48%
ELD instruction only	10	–	–	–	–	–	–	–	–	–	–	–	–
SDAIE instruction only	5	–	–	–	–	–	–	–	–	–	–	–	–
ELD instr. and SDAIE instr. but not primary language instr.	347	321	56	9%	29%	32%	22%	9%	53%	47%	50%	60%	46%
Other EL instructional services	0	–	–	–	–	–	–	–	–	–	–	–	–
None (EL only)	13	319	61	0%	46%	23%	23%	8%	49%	50%	54%	59%	39%
Program participation unknown	21	322	60	10%	29%	29%	24%	10%	54%	50%	46%	62%	46%
No special education	471	323	57	9%	27%	33%	22%	9%	55%	47%	51%	62%	46%
Special education	6	–	–	–	–	–	–	–	–	–	–	–	–
Special education unknown	0	–	–	–	–	–	–	–	–	–	–	–	–

**Table 7.B.14 Demographic Summary for RLA, Grade Eight (Target Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Number Correct Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
All population valid scores	439	323	57	9%	27%	33%	22%	9%	54%	47%	51%	62%	47%
Male	243	317	57	12%	27%	33%	20%	8%	54%	46%	50%	58%	44%
Female	194	330	55	4%	27%	33%	24%	11%	55%	48%	53%	66%	50%
Gender unknown	2	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	0	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	0	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	0	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	5	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	0	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	1	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	0	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	108	308	52	14%	30%	37%	14%	6%	48%	43%	46%	57%	40%
Guatemala	38	295	49	16%	45%	26%	11%	3%	39%	41%	44%	50%	38%
Mexico	154	339	54	3%	23%	32%	27%	14%	62%	51%	55%	68%	51%
Nicaragua	3	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	6	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	1	—	—	—	—	—	—	—	—	—	—	—	—
Spain	1	—	—	—	—	—	—	—	—	—	—	—	—
United States	95	330	59	11%	19%	34%	26%	11%	58%	48%	54%	63%	50%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	2	—	—	—	—	—	—	—	—	—	—	—	—
Other	18	290	49	11%	39%	44%	6%	0%	38%	38%	37%	50%	42%
Country unknown	7	—	—	—	—	—	—	—	—	—	—	—	—
Not economically disadvantaged	43	331	51	9%	19%	35%	28%	9%	58%	47%	57%	64%	50%
Economically disadvantaged	385	322	58	9%	28%	32%	21%	10%	54%	47%	51%	61%	46%
Economic status unknown	11	321	31	0%	27%	55%	18%	0%	45%	41%	48%	68%	52%
In U.S. schools < 12 months	416	324	57	8%	27%	33%	22%	10%	55%	47%	51%	62%	47%
In U.S. schools ≥ 12 months	23	296	42	17%	30%	39%	13%	0%	49%	33%	45%	53%	39%
Primary language instruction and ELD and/or SDAIE instruction	81	324	54	10%	20%	40%	23%	7%	58%	42%	52%	64%	48%
ELD instruction only	10	—	—	—	—	—	—	—	—	—	—	—	—
SDAIE instruction only	4	—	—	—	—	—	—	—	—	—	—	—	—
ELD instr. and SDAIE instr. but not primary language instr.	318	321	57	9%	28%	32%	21%	9%	53%	47%	50%	60%	46%
Other EL instructional services	0	—	—	—	—	—	—	—	—	—	—	—	—
None (EL only)	13	319	61	0%	46%	23%	23%	8%	49%	50%	54%	59%	39%
Program participation unknown	13	345	50	0%	15%	38%	38%	8%	61%	54%	54%	76%	51%
No special education	433	324	56	9%	26%	33%	22%	9%	55%	47%	51%	62%	47%
Special education	6	—	—	—	—	—	—	—	—	—	—	—	—
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

**Table 7.B.15 Demographic Summary for RLA, Grade Nine (Overall Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Number Correct Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
All population valid scores	1,027	319	54	9%	26%	34%	23%	8%	71%	56%	55%	53%	46%
Male	621	313	53	10%	30%	33%	20%	7%	70%	54%	52%	50%	44%
Female	403	330	53	7%	21%	34%	27%	10%	72%	60%	58%	57%	50%
Gender unknown	3	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	0	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	1	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	0	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	6	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	1	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	3	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	2	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	359	317	51	9%	29%	33%	24%	5%	71%	55%	56%	51%	45%
Guatemala	221	303	53	12%	36%	31%	15%	6%	64%	50%	49%	50%	41%
Mexico	204	337	54	5%	18%	35%	25%	17%	76%	64%	57%	59%	53%
Nicaragua	9	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	2	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	2	—	—	—	—	—	—	—	—	—	—	—	—
Spain	1	—	—	—	—	—	—	—	—	—	—	—	—
United States	103	337	52	7%	11%	39%	34%	10%	78%	63%	58%	59%	52%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	1	—	—	—	—	—	—	—	—	—	—	—	—
Other	78	304	44	9%	36%	35%	18%	3%	69%	51%	54%	46%	40%
Country unknown	34	308	57	15%	32%	24%	24%	6%	64%	53%	49%	53%	43%
Not economically disadvantaged	182	323	50	7%	24%	41%	21%	8%	73%	57%	56%	55%	47%
Economically disadvantaged	821	319	54	9%	27%	32%	23%	9%	70%	56%	55%	53%	47%
Economic status unknown	24	291	55	21%	42%	21%	17%	0%	59%	50%	48%	39%	37%
In U.S. schools < 12 months	916	320	53	8%	27%	34%	22%	9%	71%	57%	55%	53%	47%
In U.S. schools ≥ 12 months	111	310	55	17%	21%	33%	25%	4%	68%	52%	53%	52%	43%
Primary language instruction and ELD and/or SDAIE instruction	246	314	55	12%	28%	32%	19%	9%	68%	53%	52%	53%	46%
ELD instruction only	9	—	—	—	—	—	—	—	—	—	—	—	—
SDAIE instruction only	9	—	—	—	—	—	—	—	—	—	—	—	—
ELD instr. and SDAIE instr. but not primary language instr.	662	320	54	8%	26%	33%	24%	8%	71%	57%	55%	53%	46%
Other EL instructional services	11	286	45	18%	45%	36%	0%	0%	64%	48%	48%	38%	31%
None (EL only)	22	344	48	0%	14%	45%	32%	9%	78%	64%	62%	55%	58%
Program participation unknown	68	325	49	7%	18%	46%	22%	7%	75%	58%	57%	55%	48%
No special education	1,025	319	53	9%	26%	34%	23%	8%	71%	56%	55%	53%	47%
Special education	2	—	—	—	—	—	—	—	—	—	—	—	—
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

**Table 7.B.16 Demographic Summary for RLA, Grade Nine (Target Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Number Correct Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target population valid scores	956	320	54	9%	27%	33%	22%	9%	71%	56%	55%	53%	47%
Male	581	313	53	10%	30%	33%	20%	8%	69%	54%	52%	50%	45%
Female	372	330	53	6%	22%	34%	27%	11%	73%	60%	58%	57%	50%
Gender unknown	3	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	0	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	1	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	0	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	6	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	1	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	3	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	2	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	342	317	50	9%	30%	34%	23%	5%	71%	55%	56%	51%	45%
Guatemala	205	303	53	12%	36%	31%	15%	6%	64%	50%	49%	50%	41%
Mexico	185	337	56	6%	18%	34%	24%	18%	76%	64%	57%	58%	53%
Nicaragua	9	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	2	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	2	—	—	—	—	—	—	—	—	—	—	—	—
Spain	1	—	—	—	—	—	—	—	—	—	—	—	—
United States	90	343	47	2%	11%	38%	38%	11%	80%	66%	59%	61%	55%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	1	—	—	—	—	—	—	—	—	—	—	—	—
Other	75	303	44	9%	37%	33%	17%	3%	68%	50%	53%	46%	39%
Country unknown	31	308	59	16%	29%	23%	26%	6%	65%	52%	49%	53%	44%
Not economically disadvantaged	163	322	50	6%	25%	41%	20%	9%	73%	57%	56%	54%	47%
Economically disadvantaged	776	320	54	9%	27%	32%	23%	9%	70%	57%	55%	53%	47%
Economic status unknown	17	285	58	29%	35%	18%	18%	0%	57%	46%	47%	37%	36%
In U.S. schools < 12 months	916	320	53	8%	27%	34%	22%	9%	71%	57%	55%	53%	47%
In U.S. schools ≥ 12 months	40	301	61	23%	28%	20%	23%	8%	63%	48%	50%	48%	42%
Primary language instruction and ELD and/or SDAIE instruction	246	314	55	12%	28%	32%	19%	9%	68%	53%	52%	53%	46%
ELD instruction only	6	—	—	—	—	—	—	—	—	—	—	—	—
SDAIE instruction only	8	—	—	—	—	—	—	—	—	—	—	—	—
ELD instr. and SDAIE instr. but not primary language instr.	621	321	54	8%	27%	32%	24%	9%	71%	57%	55%	53%	46%
Other EL instructional services	10	—	—	—	—	—	—	—	—	—	—	—	—
None (EL only)	22	344	48	0%	14%	45%	32%	9%	78%	64%	62%	55%	58%
Program participation unknown	43	328	51	7%	16%	49%	16%	12%	74%	58%	58%	54%	51%
No special education	955	320	54	8%	27%	33%	22%	9%	71%	57%	55%	53%	47%
Special education	1	—	—	—	—	—	—	—	—	—	—	—	—
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

**Table 7.B.17 Demographic Summary for RLA, Grade Ten (Overall Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Number Correct Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
All population valid scores	407	314	58	12%	27%	32%	20%	9%	70%	55%	56%	51%	43%
Male	235	306	57	14%	30%	34%	14%	7%	67%	53%	53%	49%	42%
Female	172	325	58	10%	23%	28%	27%	11%	74%	59%	61%	55%	45%
Gender unknown	0	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	1	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	0	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	0	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	2	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	2	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	3	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	1	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	87	301	62	22%	28%	25%	18%	7%	64%	51%	54%	47%	40%
Guatemala	52	284	50	17%	46%	25%	10%	2%	60%	46%	48%	40%	35%
Mexico	130	331	53	6%	18%	40%	21%	15%	77%	61%	61%	57%	47%
Nicaragua	5	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	2	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	0	—	—	—	—	—	—	—	—	—	—	—	—
Spain	0	—	—	—	—	—	—	—	—	—	—	—	—
United States	89	313	56	12%	24%	37%	22%	4%	72%	55%	55%	53%	43%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	1	—	—	—	—	—	—	—	—	—	—	—	—
Other	16	313	55	6%	44%	31%	13%	6%	66%	55%	59%	48%	42%
Country unknown	16	289	56	13%	63%	19%	0%	6%	56%	44%	46%	44%	41%
Not economically disadvantaged	61	309	60	13%	28%	38%	15%	7%	68%	54%	55%	49%	43%
Economically disadvantaged	324	316	58	11%	27%	31%	22%	9%	71%	56%	57%	52%	44%
Economic status unknown	22	302	66	23%	36%	23%	5%	14%	64%	51%	51%	48%	41%
In U.S. schools < 12 months	338	318	57	10%	26%	33%	21%	9%	72%	56%	57%	53%	44%
In U.S. schools ≥ 12 months	69	295	60	22%	32%	29%	12%	6%	62%	50%	51%	45%	38%
Primary language instruction and ELD and/or SDAIE instruction	151	299	55	18%	30%	32%	15%	4%	64%	50%	52%	47%	40%
ELD instruction only	8	—	—	—	—	—	—	—	—	—	—	—	—
SDAIE instruction only	8	—	—	—	—	—	—	—	—	—	—	—	—
ELD instr. and SDAIE instr. but not primary language instr.	208	325	59	9%	23%	31%	25%	12%	74%	59%	60%	55%	46%
Other EL instructional services	2	—	—	—	—	—	—	—	—	—	—	—	—
None (EL only)	16	324	60	6%	38%	31%	13%	13%	78%	58%	60%	51%	44%
Program participation unknown	14	307	53	14%	29%	43%	7%	7%	68%	54%	55%	47%	41%
No special education	406	314	58	12%	27%	32%	20%	9%	70%	55%	56%	51%	43%
Special education	1	—	—	—	—	—	—	—	—	—	—	—	—
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

**Table 7.B.18 Demographic Summary for RLA, Grade Ten (Target Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Number Correct Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target population valid scores	388	314	58	12%	27%	32%	20%	9%	70%	55%	56%	52%	43%
Male	226	307	58	14%	29%	35%	15%	8%	67%	53%	53%	49%	42%
Female	162	325	56	9%	25%	29%	28%	10%	74%	59%	61%	55%	45%
Gender unknown	0	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	1	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	0	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	0	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	2	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	2	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	3	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	1	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	86	299	60	22%	28%	26%	19%	6%	64%	50%	54%	46%	39%
Guatemala	48	284	51	17%	48%	23%	10%	2%	60%	45%	49%	41%	35%
Mexico	121	332	52	6%	17%	40%	21%	15%	77%	62%	61%	57%	47%
Nicaragua	5	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	2	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	0	—	—	—	—	—	—	—	—	—	—	—	—
Spain	0	—	—	—	—	—	—	—	—	—	—	—	—
United States	86	315	54	10%	24%	38%	22%	5%	72%	56%	55%	53%	44%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	1	—	—	—	—	—	—	—	—	—	—	—	—
Other	15	315	56	7%	40%	33%	13%	7%	66%	56%	60%	49%	43%
Country unknown	15	290	58	13%	60%	20%	0%	7%	58%	45%	46%	44%	40%
Not economically disadvantaged	58	305	57	14%	28%	40%	16%	3%	67%	53%	54%	48%	42%
Economically disadvantaged	310	317	57	11%	27%	32%	22%	9%	71%	56%	57%	53%	44%
Economic status unknown	20	302	68	25%	35%	20%	5%	15%	65%	51%	51%	47%	41%
In U.S. schools < 12 months	338	318	57	10%	26%	33%	21%	9%	72%	56%	57%	53%	44%
In U.S. schools ≥ 12 months	50	291	53	22%	34%	30%	12%	2%	60%	49%	50%	44%	37%
Primary language instruction and ELD and/or SDAIE instruction	151	299	55	18%	30%	32%	15%	4%	64%	50%	52%	47%	40%
ELD instruction only	8	—	—	—	—	—	—	—	—	—	—	—	—
SDAIE instruction only	7	—	—	—	—	—	—	—	—	—	—	—	—
ELD instr. and SDAIE instr. but not primary language instr.	198	326	58	9%	23%	31%	25%	12%	74%	59%	60%	55%	46%
Other EL instructional services	1	—	—	—	—	—	—	—	—	—	—	—	—
None (EL only)	15	324	62	7%	40%	27%	13%	13%	78%	57%	60%	51%	45%
Program participation unknown	8	—	—	—	—	—	—	—	—	—	—	—	—
No special education	387	315	57	12%	27%	32%	20%	9%	70%	55%	57%	52%	43%
Special education	1	—	—	—	—	—	—	—	—	—	—	—	—
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

**Table 7.B.19 Demographic Summary for RLA, Grade Eleven (Overall Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Number Correct Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
All population valid scores	198	319	61	11%	23%	30%	26%	11%	64%	50%	56%	53%	56%
Male	106	311	63	13%	26%	28%	20%	12%	60%	48%	52%	50%	54%
Female	91	329	57	7%	19%	33%	33%	9%	68%	51%	60%	56%	57%
Gender unknown	1	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	2	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	0	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	0	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	4	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	1	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	0	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	0	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	34	297	63	18%	29%	29%	18%	6%	58%	48%	49%	46%	47%
Guatemala	19	285	59	26%	16%	47%	11%	0%	52%	42%	45%	46%	46%
Mexico	87	334	51	2%	24%	30%	31%	13%	68%	53%	60%	56%	61%
Nicaragua	3	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	3	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	0	—	—	—	—	—	—	—	—	—	—	—	—
Spain	2	—	—	—	—	—	—	—	—	—	—	—	—
United States	33	311	61	9%	27%	30%	27%	6%	59%	44%	55%	49%	56%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	1	—	—	—	—	—	—	—	—	—	—	—	—
Other	4	—	—	—	—	—	—	—	—	—	—	—	—
Country unknown	5	—	—	—	—	—	—	—	—	—	—	—	—
Not economically disadvantaged	29	340	60	7%	10%	24%	45%	14%	70%	54%	62%	59%	63%
Economically disadvantaged	160	316	61	11%	24%	32%	23%	11%	63%	49%	55%	52%	55%
Economic status unknown	9	—	—	—	—	—	—	—	—	—	—	—	—
In U.S. schools < 12 months	161	328	57	8%	19%	31%	31%	11%	67%	51%	59%	55%	58%
In U.S. schools ≥ 12 months	37	280	63	22%	41%	27%	3%	8%	51%	45%	41%	42%	43%
Primary language instruction and ELD and/or SDAIE instruction	56	298	59	14%	32%	29%	21%	4%	58%	45%	51%	44%	49%
ELD instruction only	6	—	—	—	—	—	—	—	—	—	—	—	—
SDAIE instruction only	5	—	—	—	—	—	—	—	—	—	—	—	—
ELD instr. and SDAIE instr. but not primary language instr.	114	330	61	8%	20%	30%	27%	15%	67%	52%	58%	57%	59%
Other EL instructional services	1	—	—	—	—	—	—	—	—	—	—	—	—
None (EL only)	6	—	—	—	—	—	—	—	—	—	—	—	—
Program participation unknown	10	—	—	—	—	—	—	—	—	—	—	—	—
No special education	197	320	60	10%	23%	30%	26%	11%	64%	50%	56%	53%	56%
Special education	1	—	—	—	—	—	—	—	—	—	—	—	—
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

**Table 7.B.20 Demographic Summary for RLA, Grade Eleven (Target Population)**

	Number Tested	Mean Scale Score	Std. Dev. of Number Correct Scores	Percent in Performance Level					Mean Percent Correct in Reporting Cluster				
				Far Below Basic	Below Basic	Basic	Proficient	Advanced	Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target population valid scores	183	319	61	11%	21%	30%	28%	10%	64%	49%	56%	53%	56%
Male	99	311	64	14%	24%	28%	21%	12%	60%	48%	53%	51%	54%
Female	83	330	55	6%	18%	33%	36%	7%	68%	51%	61%	56%	58%
Gender unknown	1	—	—	—	—	—	—	—	—	—	—	—	—
Argentina	2	—	—	—	—	—	—	—	—	—	—	—	—
Bolivia	0	—	—	—	—	—	—	—	—	—	—	—	—
Brazil	0	—	—	—	—	—	—	—	—	—	—	—	—
Chile	0	—	—	—	—	—	—	—	—	—	—	—	—
Colombia	4	—	—	—	—	—	—	—	—	—	—	—	—
Costa Rica	1	—	—	—	—	—	—	—	—	—	—	—	—
Cuba	0	—	—	—	—	—	—	—	—	—	—	—	—
Ecuador	0	—	—	—	—	—	—	—	—	—	—	—	—
El Salvador	29	293	65	21%	28%	28%	21%	3%	57%	47%	48%	44%	46%
Guatemala	19	285	59	26%	16%	47%	11%	0%	52%	42%	45%	46%	46%
Mexico	80	334	50	3%	23%	30%	34%	11%	69%	52%	61%	57%	60%
Nicaragua	3	—	—	—	—	—	—	—	—	—	—	—	—
Panama	0	—	—	—	—	—	—	—	—	—	—	—	—
Paraguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Peru	3	—	—	—	—	—	—	—	—	—	—	—	—
Puerto Rico	0	—	—	—	—	—	—	—	—	—	—	—	—
Spain	2	—	—	—	—	—	—	—	—	—	—	—	—
United States	31	317	56	6%	26%	32%	29%	6%	60%	45%	57%	51%	58%
Uruguay	0	—	—	—	—	—	—	—	—	—	—	—	—
Venezuela	1	—	—	—	—	—	—	—	—	—	—	—	—
Other	4	—	—	—	—	—	—	—	—	—	—	—	—
Country unknown	4	—	—	—	—	—	—	—	—	—	—	—	—
Not economically disadvantaged	28	347	51	4%	11%	25%	46%	14%	71%	55%	64%	60%	65%
Economically disadvantaged	148	316	62	12%	22%	32%	24%	9%	63%	48%	55%	52%	54%
Economic status unknown	7	—	—	—	—	—	—	—	—	—	—	—	—
In U.S. schools < 12 months	161	328	57	8%	19%	31%	31%	11%	67%	51%	59%	55%	58%
In U.S. schools ≥ 12 months	22	257	50	32%	41%	23%	5%	0%	42%	39%	37%	39%	34%
Primary language instruction and ELD and/or SDAIE instruction	56	298	59	14%	32%	29%	21%	4%	58%	45%	51%	44%	49%
ELD instruction only	6	—	—	—	—	—	—	—	—	—	—	—	—
SDAIE instruction only	5	—	—	—	—	—	—	—	—	—	—	—	—
ELD instr. and SDAIE instr. but not primary language instr.	103	331	59	8%	18%	30%	30%	14%	67%	52%	59%	58%	59%
Other EL instructional services	1	—	—	—	—	—	—	—	—	—	—	—	—
None (EL only)	5	—	—	—	—	—	—	—	—	—	—	—	—
Program participation unknown	7	—	—	—	—	—	—	—	—	—	—	—	—
No special education	183	319	61	11%	21%	30%	28%	10%	64%	49%	56%	53%	56%
Special education	0	—	—	—	—	—	—	—	—	—	—	—	—
Special education unknown	0	—	—	—	—	—	—	—	—	—	—	—	—

## Appendix 7.C—Types of Score Reports

**Table 7.C.1 Score Reports Reflecting STS Results**

<b>2014 CAASPP STS Printed Reports</b>	
<b>Description</b>	<b>Distribution</b>
<b>The STS Student Report</b>	
<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"> <li>• Scale scores</li> <li>• Performance levels (advanced, proficient, basic, below basic, and far below basic)</li> <li>• Number and percent correct in each reporting cluster</li> </ul> <p>Because students who take the grade-level STS must also take the required CSTs or CMA for Science in grade five, eight, or ten, those students will likely receive two Student Reports.</p>	<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 distributed by the LEA to parents/guardians.</p>
<b>Student Record Label</b>	
<p>These reports are printed on adhesive labels to be affixed to the student’s permanent school records. Each student shall have an individual record of accomplishment that includes CAASPP testing results (see California <i>EC</i> Section 60607[a]). Data presented include the following for each content area:</p> <ul style="list-style-type: none"> <li>• Scale scores</li> <li>• Performance levels (advanced, proficient, basic, below basic, and far below basic)</li> </ul>	<p>This report includes individual student results and is not distributed beyond the student’s school.</p>
<b>Student Master List</b>	
<p>This report is an alphabetical roster that presents individual student results. The following results are provided:</p> <ul style="list-style-type: none"> <li>• Percent correct for each reporting cluster within each content area tested</li> <li>• A scale score and a performance level for the grade-level test or tests taken</li> </ul>	<p>This report provides administrators and teachers with all students’ results within each grade or within each grade and year-round schedule at a school.</p> <p>Because this report includes individual student results, it is not distributed beyond the student’s school. It is recommended that Student Master List reports be retained until the grade level exits the school.</p>

<b>2014 CAASPP STS Printed Reports</b>	
<b>Description</b>	<b>Distribution</b>
<b>Student Master List Summary</b>	
<p>This report summarizes student results at the school, LEA, county, and state levels for each grade. It does not include any individual student information.</p> <p>For each STS, the following data are summarized:</p> <ul style="list-style-type: none"> <li>• By content area tested: <ul style="list-style-type: none"> <li>– Number of students enrolled</li> <li>– Number and percent of students tested</li> <li>– Number and percent of valid scores</li> <li>– Number tested with scores</li> <li>– Mean percent correct</li> </ul> </li> <li>• Mean scale score</li> <li>• Scale score standard deviation</li> <li>• Number and percent of students scoring at each performance level</li> <li>• The number of items for each reporting cluster and the mean percent correct</li> </ul>	<p>This report is a resource for evaluators, researchers, teachers, parents/guardians, community members, and administrators.</p> <p>One copy is packaged for the school and one for the LEA.</p> <p>This report is also produced for LEAs, counties, and the state.</p> <p><b>Note:</b> The data in this report may be shared with parents/guardians, community members, and the media only if the data are for 11 or more students. It is recommended that summary reports be retained until the grade level exits the school.</p>
<b>Subgroup Summary</b>	
<p>This set of reports disaggregates and reports results by the following subgroups:</p> <ul style="list-style-type: none"> <li>• All students</li> <li>• Disability status</li> <li>• Economic status</li> <li>• Gender</li> <li>• English proficiency</li> <li>• Primary ethnicity</li> </ul> <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"> <li>• Total number tested in the subgroup</li> <li>• Percent of enrollment tested in the subgroup</li> <li>• Number and percent of valid scores</li> <li>• Number tested who received scores</li> <li>• Mean scale score</li> <li>• Scale score standard deviation</li> <li>• Number and percent of students scoring at each performance level</li> </ul>	<p>This report is a resource for evaluators, researchers, teachers, parents/guardians, community members, and administrators.</p> <p>One copy is packaged for the school and one for the LEA.</p> <p>This report is also produced for LEAs, counties, and the state.</p> <p><b>Note:</b> 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. It is recommended that summary reports be retained until the grade level exits the school.</p>

## Chapter 8: Analyses

---

This chapter summarizes the item- and test-level statistics obtained for the STS administered during the spring of 2014 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 2014, 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, because the intact or modified intact forms were used, DIF analyses were not performed.

Each of the sets of analyses is presented in the body of the text and in the appendices as listed below.

1. Appendix 8.A on page 153 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 137.
2. Appendix 8.B on page 159 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 175 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 analyses were conducted at different times after test administration and involved varying proportions of the full STS 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 target population using the P1 data file, which contained the entire test-taking population and all the student records used in the reporting of CAASPP results.

During the 2014 administration, neither IRT calibrations nor scaling are implemented because of the intact forms and pre-equating. For the intact forms used without any replacement or edited items, 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 145 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:

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

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 153. Each set of values is presented for both the current and the original presentation of each STS.

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

Content Area	STS *	No. of Items	No. of Examinees	Mean		Median	
				$p$ -value	Pt-Bis	$p$ -value	Pt-Bis
Reading/ Language Arts	2	65	1,930	0.59	0.44	0.58	0.44
	3	65	1,388	0.50	0.36	0.48	0.35
	4	75	803	0.55	0.40	0.56	0.42
	5	75	645	0.49	0.37	0.48	0.38
	6	75	482	0.50	0.38	0.51	0.39
	7	75	460	0.53	0.37	0.53	0.40
	8	75	439	0.52	0.37	0.51	0.37
	9	75	956	0.54	0.37	0.52	0.38
	10	75	388	0.53	0.37	0.54	0.38
	11	75	183	0.55	0.35	0.55	0.38

\* STS named by number only are grade-level tests.

## 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 may 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,

$\alpha$  is the reliability estimated in equation 8.3, and

$s_t$  is the standard deviation 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  $\pm 1.96$  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 ( $15 \pm 3.76$  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**

Content Area	STS *	Test Length	No. Examinees	Reliab.	Scale Score			Raw Score		
					Mean	SD	SEM	Mean	SD	SEM
Reading/ Language Arts	2	65	1,930	0.93	319	55	14.28	38.31	13.49	3.49
	3	65	1,388	0.89	312	45	14.57	32.78	11.21	3.66
	4	75	803	0.93	316	52	13.94	41.17	14.34	3.86
	5	75	645	0.92	313	59	16.96	36.80	13.75	3.95
	6	75	482	0.92	318	62	17.63	37.46	13.81	3.95
	7	75	460	0.92	323	57	16.42	39.59	13.54	3.88
	8	75	439	0.92	323	57	16.31	38.77	13.56	3.91
	9	75	956	0.91	320	54	15.94	40.79	12.89	3.82
	10	75	388	0.91	314	58	17.12	40.01	13.02	3.87
	11	75	183	0.90	319	61	19.03	41.02	12.57	3.93

\* STS named by number only are grade-level tests.

### 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, length of attendance in U.S. schools, and EL program participation.

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.6 present the overall test reliabilities for the various subgroups. Table 8.B.7 presents the cluster-level reliabilities for the subgroups based on gender and economic status. The next table, Table 8.B.8, shows the same analyses for the subgroups based on provision of special services and the length of attendance in U.S. schools. The last table, Table 8.B.9, present the cluster-based reliabilities for the subgroups based on EL program participation.

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 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,

$CSEM(\hat{\theta})$  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.

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**

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

\* STS named by number only are grade-level tests.

## 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 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.10 through Table 8.B.19 in Appendix 8.B, starting on page 170.

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], 1999; 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, 1999; 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 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 (USDOE) for the peer review of testing programs administered by states in response to the Elementary and Secondary Education Act (USDOE, 2001).

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

### **Purpose of the STS**

As mentioned in Chapter 1, the purpose of the STS program is to permit students to demonstrate achievement of the California content standards in reading/language arts 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 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 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 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, student performance levels, and subscores for each reporting cluster 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. Reporting cluster results compare 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 is also provided for each cluster using a percent-correct metric. The reference points for this range are: (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 scores is presented in Chapter 7, which starts on page 102.

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 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, 1999, Standard 1.3). The user is advised to gather evidence to support these additional interpretations or uses (AERA, APA, & NCME, 1999, Standard 1.4).

## Intended Test Population(s)

The STS are optional tests that are targeted toward Spanish-speaking 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. 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.

## Validity Evidence Collected

### Evidence Based on Content

According to *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), 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* (1999), 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: ELA, 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.

**Specifications and blueprints**—ETS maintains item specifications for each STS. 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 58. Once the items are developed and field-tested, ETS selects all STS test items to conform to the SBE-approved California content standards and test blueprints. Test blueprints for the STS 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. 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. The test blueprints for the STS 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 is presented in Chapter 3, starting on page 58.

**Item review process**—Chapter 3 explains in detail the extensive item review process applied to items written for use in the STS. In brief, items written for the STS 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 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 68.

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 64.

### **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, 1999), 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, DIF analyses are 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’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 STS for RLA is shown in Table 8.4.

**Table 8.4 Original Year of Administration for the STS**

Content Area	STS *	Year
	2	2012
	3	2011
	4	2012
	5	2011
Reading/Language Arts	6	2012
	7	2011
	8	2012
	9	2011
	10	2012
	11	2011

\* STS named by number only are grade-level tests.

### **Evidence Based on Response Processes**

As noted in the APA, AERA, and NCME *Standards* (1999), 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, 1999), 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, 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 0.35 and 0.44. The point biserials for the individual items in the STS 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 in the STS for RLA in grades six and eight, which included replacement items, can be found in Table 8.C.1 and Table 8.C.2, respectively. The distributions of the  $b$ -values for these tests that contained replacement items are given in Table 8.C.3. For operational items in the other tests, the summaries of  $b$ -values 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 145 for administration years.)

The data in Table 8.1 indicate that STS tests had average  $p$ -values that range from 0.49 to 0.59.

### **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 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 0.89 to 0.93.

**Reporting cluster reliabilities**—For each STS, 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 the results.

**Subgroup reliabilities**—The reliabilities of the operational STS 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.6.

**Reliability of performance classifications**—The methodology used for estimating the reliability of classification decisions is described in the section “Decision Classification Analyses” on page 140. The results of these analyses are presented in Table 8.B.10 through Table 8.B.19 in Appendix 8.B; these tables start on page 170. When the classifications are collapsed to below-proficient versus proficient and above, the proportion of students that were classified accurately ranged from 0.88 to 0.94 across all the STS. Similarly, the proportion of students that were classified consistently ranged from 0.85 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.

## IRT Analyses

### Post-equating

Prior to 2014, the STS 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 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 prior to 2014 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 tests.

### Pre-Equating

During the 2014 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 item response theory (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 intact forms without any edits or replacement of items, the conversion tables from previous administrations when the forms were originally used are directly applied to the current administration. Otherwise, in the cases where replacement or edited items are used, conversion tables are generated using the true-score equating method described in Chapter 2, on page 15. The item parameters used for true-score equating are post-equating item parameters from the intact forms for the unchanged items and the post-equating item parameters from the most recent administration for the replacement or edited items.

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 is shown in Table 8.4

The STS for RLA in grades six and eight have IRT  $b$ -value and distribution tables available in Appendix 8.C starting on page 174 because they contain six replacement items.

The details on all equating procedures are given in Chapter 2, starting on page 13.

### Summaries of Scaled IRT $b$ -values

For the post-equating procedure prior to the 2014 administration—the first administration in which forms were reused—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 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. 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 145 for administration years.

Table 8.C.1 and Table 8.C.2 present univariate statistics (mean, standard deviation, minimum, and maximum) for the scaled IRT  $b$ -values for the tests containing replacement or edited items. The results for the overall test are presented for the operational items in each reporting cluster. (For operational items of the other tests, the summaries of  $b$ -values can be found in Appendix D of the *STS Technical Report* in the year each grade-level RLA form was administered originally.)

Table 8.C.3 shows the distributions of operational items for tests containing replacement items across 16 intervals of  $b$ -values. In these tables, the intervals range from “greater than

or equal to 3.5” to “less than –3.5” points within each interval. (For operational items of the other tests, the distributions 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 145 for administration years.)

### Evaluation of Pre-equating

Pre-equating is performed on the basis of the assumption of item response theory (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 administration, for all STS intact forms, the conversion tables from their original administrations (listed in Table 8.4 on page 145) are directly applied to the current administration.

Complete raw-score-to-scale-score conversion tables for the STS administered in 2014 are presented in Table 8.C.4 through Table 8.C.13 starting on page 184. 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 2014 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 administration because forms are either intact or intact with some modification, 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 145 for administration years.

The statistical procedure of DIF analysis that was conducted prior to the 2014 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  $\alpha_{MH}$  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"> <li>• Absolute value of MH D-DIF is not significantly different from zero, or is less than one.</li> <li>• Positive values are classified as “A+” and negative values as “A-.”</li> </ul>
B (moderate)	<ul style="list-style-type: none"> <li>• Absolute value of MH D-DIF is significantly different from zero but not from one, and is at least one; OR</li> <li>• Absolute value of MH D-DIF is significantly different from one, but is less than 1.5.</li> <li>• Positive values are classified as “B+” and negative values as “B-.”</li> </ul>
C (large)	<ul style="list-style-type: none"> <li>• Absolute value of MH D-DIF is significantly different from one, and is at least 1.5.</li> <li>• Positive values are classified as “C+” and negative values as “C-.”</li> </ul>

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. 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. 1999. *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.htm>

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**

Item-by-item $p$ -value and Point Biserial for RLA								
STS	2				3			
Year	2012		2014		2011		2014	
Items	$p$ -value	Pt-Bis						
1	0.91	0.29	0.85	0.33	0.49	0.36	0.25	0.14
2	0.76	0.49	0.71	0.47	0.92	0.26	0.87	0.29
3	0.31	0.25	0.31	0.25	0.48	0.26	0.46	0.17
4	0.73	0.45	0.70	0.44	0.70	0.43	0.62	0.34
5	0.71	0.31	0.58	0.31	0.31	0.28	0.23	0.22
6	0.60	0.33	0.56	0.32	0.63	0.47	0.58	0.47
7	0.82	0.46	0.73	0.45	0.59	0.41	0.57	0.36
8	0.76	0.52	0.74	0.51	0.75	0.39	0.67	0.30
9	0.67	0.43	0.65	0.42	0.87	0.42	0.80	0.44
10	0.80	0.53	0.76	0.52	0.71	0.48	0.63	0.50
11	0.47	0.36	0.40	0.32	0.52	0.39	0.44	0.35
12	0.91	0.42	0.88	0.47	0.66	0.38	0.54	0.35
13	0.62	0.48	0.58	0.44	0.53	0.31	0.46	0.29
14	0.63	0.54	0.59	0.56	0.80	0.37	0.73	0.35
15	0.64	0.53	0.59	0.46	0.53	0.44	0.48	0.43
16	0.62	0.47	0.62	0.43	0.32	0.20	0.31	0.25
17	0.68	0.38	0.62	0.41	0.58	0.44	0.51	0.46
18	0.78	0.35	0.74	0.34	0.54	0.34	0.45	0.36
19	0.66	0.47	0.59	0.43	0.45	0.30	0.40	0.28
20	0.73	0.57	0.68	0.55	0.74	0.40	0.66	0.41
21	0.80	0.45	0.78	0.43	0.59	0.48	0.52	0.48
22	0.62	0.45	0.62	0.49	0.56	0.45	0.51	0.43
23	0.56	0.50	0.48	0.46	0.62	0.46	0.59	0.39
24	0.59	0.47	0.57	0.43	0.77	0.47	0.72	0.48
25	0.60	0.47	0.52	0.42	0.86	0.35	0.81	0.35
26	0.66	0.45	0.62	0.45	0.76	0.37	0.70	0.43
27	0.63	0.47	0.61	0.47	0.53	0.37	0.45	0.31
28	0.43	0.27	0.42	0.31	0.63	0.44	0.46	0.30
29	0.79	0.51	0.75	0.53	0.76	0.28	0.68	0.30
30	0.81	0.49	0.77	0.49	0.51	0.39	0.46	0.32
31	0.59	0.49	0.48	0.43	0.57	0.38	0.53	0.38
32	0.63	0.47	0.58	0.46	0.58	0.38	0.53	0.39
33	0.64	0.44	0.60	0.46	0.35	0.32	0.33	0.32
34	0.48	0.46	0.43	0.41	0.33	0.23	0.31	0.17
35	0.56	0.37	0.58	0.40	0.47	0.44	0.40	0.43
36	0.59	0.44	0.52	0.42	0.68	0.45	0.61	0.44
37	0.80	0.50	0.74	0.52	0.37	0.32	0.36	0.26
38	0.51	0.44	0.44	0.47	0.39	0.28	0.35	0.25
39	0.56	0.43	0.52	0.38	0.53	0.44	0.45	0.40
40	0.55	0.44	0.49	0.46	0.32	0.24	0.30	0.24
41	0.80	0.40	0.75	0.42	0.60	0.40	0.44	0.35
42	0.79	0.54	0.74	0.56	0.68	0.49	0.62	0.47
43	0.73	0.53	0.67	0.54	0.67	0.46	0.59	0.46
44	0.69	0.54	0.62	0.53	0.57	0.41	0.42	0.38
45	0.59	0.48	0.47	0.45	0.50	0.42	0.40	0.33
46	0.62	0.44	0.57	0.43	0.61	0.39	0.47	0.34
47	0.42	0.29	0.42	0.30	0.64	0.35	0.45	0.31
48	0.46	0.34	0.41	0.30	0.59	0.53	0.44	0.46

Item-by-item $p$ -value and Point Biserial for RLA								
STS	2				3			
Year	2012		2014		2011		2014	
Items	$p$ -value	Pt-Bis						
49	0.58	0.47	0.49	0.44	0.54	0.32	0.54	0.30
50	0.77	0.55	0.72	0.56	0.62	0.46	0.55	0.42
51	0.63	0.42	0.57	0.38	0.51	0.35	0.48	0.34
52	0.48	0.35	0.47	0.30	0.50	0.49	0.38	0.47
53	0.44	0.47	0.38	0.39	0.72	0.44	0.56	0.41
54	0.73	0.53	0.65	0.51	0.65	0.42	0.53	0.39
55	0.62	0.49	0.51	0.44	0.53	0.40	0.39	0.35
56	0.55	0.46	0.49	0.42	0.76	0.53	0.66	0.54
57	0.59	0.46	0.54	0.43	0.58	0.31	0.53	0.23
58	0.68	0.54	0.63	0.56	0.51	0.39	0.39	0.31
59	0.59	0.40	0.53	0.40	0.46	0.35	0.39	0.27
60	0.69	0.46	0.66	0.48	0.54	0.41	0.50	0.34
61	0.61	0.55	0.56	0.55	0.57	0.53	0.48	0.52
62	0.62	0.54	0.50	0.49	0.73	0.56	0.59	0.53
63	0.46	0.38	0.44	0.36	0.58	0.37	0.51	0.31
64	0.59	0.51	0.57	0.45	0.38	0.21	0.32	0.16

**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**

Item-by-item $p$ -value and Point Biserial for RLA												
STS	4				5				6			
Year	2012		2014		2011		2014		2012		2014	
Items	$p$ -value	Pt-Bis										
1	0.52	0.45	0.52	0.42	0.60	0.44	0.60	0.51	0.62	0.33	0.59	0.42
2	0.69	0.43	0.66	0.40	0.51	0.30	0.44	0.27	0.42	0.23	0.45	0.17
3	0.70	0.46	0.71	0.43	0.45	0.26	0.51	0.29	0.65	0.40	0.62	0.44
4	0.61	0.40	0.47	0.36	0.36	0.23	0.40	0.21	0.68	0.35	0.60	0.40
5	0.45	0.36	0.42	0.35	0.74	0.36	0.73	0.33	0.65	0.33	0.58	0.37
6	0.67	0.52	0.66	0.53	0.54	0.46	0.56	0.48	0.63	0.40	0.56	0.41
7	0.80	0.46	0.73	0.54	0.69	0.50	0.68	0.49	0.49	0.33	0.42	0.46
8	0.77	0.35	0.77	0.35	0.53	0.48	0.43	0.44	0.47	0.34	0.47	0.42
9	0.70	0.43	0.60	0.30	0.65	0.46	0.54	0.38	0.41	0.39	0.35	0.42
10	0.69	0.50	0.63	0.45	0.45	0.29	0.42	0.34	0.59	0.35	0.62	0.42
11	0.70	0.44	0.67	0.42	0.49	0.37	0.47	0.43	0.65	0.45	0.65	0.41
12	0.62	0.37	0.63	0.32	0.60	0.44	0.62	0.43	0.58	0.45	0.61	0.47
13	0.69	0.46	0.64	0.51	0.50	0.36	0.49	0.41	0.59	0.49	0.55	0.49
14	0.60	0.46	0.48	0.36	0.45	0.43	0.41	0.34	0.70	0.36	0.68	0.30
15	0.52	0.35	0.53	0.36	0.58	0.40	0.55	0.35	0.49	0.26	0.51	0.22
16	0.51	0.31	0.55	0.28	0.39	0.21	0.38	0.18	0.52	0.34	0.51	0.39
17	0.73	0.48	0.64	0.51	0.49	0.48	0.46	0.50	0.67	0.46	0.63	0.47
18	0.78	0.48	0.76	0.46	0.45	0.42	0.48	0.43	0.66	0.46	0.65	0.37
19	0.65	0.42	0.65	0.41	0.58	0.54	0.60	0.56	0.53	0.44	0.46	0.44
20	0.54	0.40	0.50	0.36	0.34	0.31	0.36	0.32	0.52	0.40	0.55	0.39
21	0.51	0.35	0.48	0.32	0.45	0.34	0.43	0.33	0.54	0.43	0.54	0.39
22	0.55	0.47	0.55	0.44	0.43	0.34	0.37	0.31	0.36	0.25	0.34	0.27
23	0.49	0.32	0.50	0.31	0.44	0.50	0.44	0.49	0.42	0.41	0.40	0.41
24	0.36	0.27	0.33	0.18	0.40	0.42	0.38	0.33	0.53	0.37	0.56	0.29
25	0.50	0.42	0.47	0.45	0.54	0.47	0.60	0.44	0.61	0.40	0.55	0.42
26	0.55	0.40	0.50	0.39	0.42	0.35	0.45	0.35	0.44	0.32	0.41	0.31
27	0.62	0.52	0.53	0.44	0.35	0.21	0.35	0.19	0.56	0.42	0.52	0.50
28	0.53	0.38	0.50	0.36	0.37	0.31	0.37	0.32	0.42	0.24	0.44	0.23
29	0.68	0.54	0.63	0.55	0.39	0.33	0.39	0.29	0.56	0.48	0.59	0.50

Item-by-item $p$ -value and Point Biserial for RLA												
STS	4				5				6			
Year	2012		2014		2011		2014		2012		2014	
Items	$p$ -value	Pt-Bis										
30	0.71	0.53	0.67	0.59	0.56	0.41	0.53	0.46	0.45	0.27	0.44	0.25
31	0.65	0.54	0.57	0.45	0.42	0.39	0.42	0.45	0.51	0.36	0.50	0.27
32	0.63	0.39	0.65	0.30	0.41	0.34	0.37	0.31	0.30	0.22	0.25	0.24
33	0.22	0.06	0.25	0.06	0.53	0.42	0.49	0.44	0.43	0.35	0.43	0.29
34	0.56	0.39	0.61	0.39	0.40	0.27	0.38	0.24	0.45	0.28	0.43	0.24
35	0.39	0.22	0.39	0.21	0.37	0.32	0.35	0.30	0.37	0.27	0.39	0.24
36	0.58	0.26	0.56	0.26	0.49	0.45	0.50	0.47	0.26	0.13	0.21	0.09
37	0.86	0.42	0.75	0.51	0.49	0.46	0.52	0.50	0.41	0.29	0.40	0.29
38	0.88	0.45	0.84	0.50	0.58	0.40	0.54	0.32	0.63	0.39	0.65	0.37
39	0.76	0.46	0.64	0.54	0.66	0.47	0.63	0.45	0.46	0.36	0.50	0.36
40	0.66	0.41	0.57	0.43	0.52	0.36	0.42	0.38	0.49	0.51	0.45	0.53
41	0.74	0.40	0.68	0.49	0.83	0.39	0.77	0.42	0.67	0.31	0.63	0.32
42	0.78	0.41	0.67	0.47	0.72	0.29	0.68	0.38	0.56	0.42	0.53	0.42
43	0.51	0.35	0.43	0.35	0.74	0.44	0.71	0.51	0.48	0.40	0.45	0.33
44	0.70	0.50	0.62	0.57	0.63	0.45	0.62	0.39	0.51	0.38	0.46	0.45
45	0.62	0.45	0.57	0.52	0.55	0.39	0.48	0.40	0.47	0.31	0.47	0.36
46	0.49	0.37	0.45	0.34	0.54	0.32	0.50	0.35	0.41	0.30	0.39	0.29
47	0.32	0.27	0.25	0.22	0.73	0.44	0.67	0.52	0.86	0.43	0.77	0.52
48	0.60	0.42	0.56	0.47	0.46	0.32	0.43	0.37	0.52	0.33	0.46	0.38
49	0.60	0.39	0.54	0.37	0.53	0.28	0.49	0.27	0.36	0.38	0.33	0.42
50	0.35	0.15	0.33	0.17	0.57	0.32	0.56	0.36	0.62	0.42	0.61	0.43
51	0.37	0.30	0.32	0.26	0.51	0.42	0.49	0.40	0.51	0.32	0.48	0.35
52	0.56	0.45	0.53	0.43	0.46	0.33	0.41	0.35	0.66	0.51	0.63	0.56
53	0.69	0.51	0.65	0.54	0.54	0.33	0.46	0.30	0.47	0.32	0.43	0.34
54	0.70	0.44	0.66	0.46	0.70	0.42	0.69	0.42	0.57	0.38	0.59	0.36
55	0.52	0.52	0.48	0.55	0.32	0.24	0.30	0.17	0.55	0.42	0.54	0.39
56	0.52	0.44	0.51	0.44	0.54	0.34	0.53	0.40	0.52	0.36	0.50	0.38
57	0.54	0.44	0.52	0.43	0.63	0.38	0.60	0.38	0.45	0.29	0.35	0.26
58	0.63	0.48	0.50	0.48	0.51	0.28	0.47	0.22	0.52	0.52	0.51	0.54
59	0.36	0.31	0.33	0.22	0.45	0.35	0.48	0.36	0.40	0.36	0.38	0.32
60	0.58	0.28	0.52	0.22	0.50	0.44	0.52	0.48	0.40	0.29	0.37	0.26
61	0.36	0.28	0.28	0.23	0.38	0.26	0.37	0.25	0.45	0.35	0.43	0.33
62	0.60	0.44	0.58	0.43	0.45	0.28	0.43	0.23	0.65	0.50	0.65	0.51
63	0.71	0.48	0.65	0.50	0.63	0.47	0.58	0.50	0.55	0.38	0.52	0.44
64	0.41	0.36	0.42	0.35	0.58	0.30	0.62	0.26	0.63	0.45	0.59	0.52
65	0.72	0.48	0.65	0.55	0.59	0.51	0.57	0.51	0.42	0.32	0.50	0.34
66	0.36	0.28	0.36	0.30	0.50	0.35	0.48	0.32	0.64	0.46	0.55	0.52
67	0.56	0.44	0.51	0.30	0.51	0.39	0.50	0.38	0.70	0.41	0.59	0.40
68	0.50	0.42	0.47	0.42	0.39	0.50	0.39	0.57	0.49	0.34	0.51	0.39
69	0.62	0.42	0.57	0.37	0.51	0.38	0.51	0.38	0.58	0.51	0.61	0.49
70	0.64	0.43	0.62	0.43	0.29	0.32	0.26	0.22	0.66	0.48	0.60	0.57
71	0.63	0.46	0.61	0.48	0.50	0.47	0.49	0.49	0.32	0.23	0.28	0.15
72	0.67	0.45	0.62	0.48	0.40	0.36	0.31	0.35	0.57	0.51	0.56	0.49
73	0.71	0.53	0.69	0.50	0.46	0.39	0.44	0.40	0.39	0.23	0.35	0.26
74	0.42	0.33	0.41	0.31	0.32	0.32	0.32	0.30	0.50	0.35	0.49	0.37
75	0.49	0.31	0.43	0.25	0.61	0.47	0.58	0.46	0.32	0.29	0.31	0.29

**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**

Item-by-item $p$ -value and Point Biserial for RLA												
STS	7				8				9			
Year	2011		2014		2012		2014		2011		2014	
Items	$p$ -value	Pt-Bis										
1	0.89	0.17	0.86	0.21	0.52	0.36	0.50	0.33	0.29	0.19	0.27	0.26
2	0.58	0.25	0.61	0.35	0.48	0.42	0.41	0.27	0.37	0.22	0.37	0.27
3	0.73	0.42	0.76	0.38	0.77	0.41	0.40	0.29	0.72	0.25	0.70	0.19
4	0.74	0.31	0.76	0.26	0.65	0.33	0.62	0.26	0.37	0.30	0.36	0.23
5	0.51	0.30	0.51	0.37	0.51	0.29	0.44	0.32	0.43	0.30	0.38	0.24
6	0.60	0.41	0.60	0.44	0.48	0.32	0.47	0.26	0.27	0.36	0.19	0.33
7	0.76	0.39	0.71	0.44	0.53	0.46	0.54	0.52	0.84	0.35	0.83	0.34
8	0.31	0.39	0.27	0.37	0.73	0.39	0.70	0.40	0.75	0.39	0.77	0.41
9	0.78	0.40	0.72	0.41	0.74	0.44	0.67	0.50	0.61	0.30	0.55	0.30
10	0.76	0.43	0.73	0.45	0.48	0.37	0.51	0.26	0.43	0.15	0.49	0.18
11	0.45	0.43	0.43	0.44	0.49	0.23	0.46	0.25	0.83	0.45	0.81	0.48
12	0.46	0.29	0.38	0.37	0.72	0.30	0.65	0.32	0.76	0.44	0.73	0.44
13	0.70	0.33	0.65	0.38	0.70	0.31	0.64	0.31	0.82	0.51	0.80	0.50
14	0.38	0.40	0.36	0.44	0.58	0.25	0.53	0.28	0.49	0.26	0.49	0.32
15	0.57	0.31	0.58	0.30	0.45	0.26	0.49	0.31	0.47	0.33	0.44	0.30
16	0.50	0.47	0.39	0.51	0.42	0.38	0.41	0.47	0.41	0.28	0.38	0.27
17	0.55	0.32	0.54	0.36	0.53	0.37	0.53	0.44	0.80	0.44	0.77	0.45
18	0.73	0.45	0.75	0.45	0.56	0.49	0.59	0.44	0.70	0.47	0.74	0.43
19	0.47	0.43	0.45	0.42	0.34	0.21	0.39	0.24	0.66	0.33	0.68	0.40
20	0.41	0.13	0.40	0.12	0.74	0.43	0.72	0.46	0.72	0.55	0.73	0.49
21	0.44	0.22	0.53	0.28	0.42	0.48	0.39	0.48	0.61	0.36	0.57	0.33
22	0.57	0.44	0.54	0.41	0.63	0.49	0.61	0.42	0.83	0.51	0.84	0.43
23	0.50	0.27	0.50	0.43	0.46	0.29	0.43	0.34	0.43	0.27	0.42	0.32
24	0.35	0.23	0.37	0.28	0.37	0.38	0.36	0.37	0.80	0.46	0.82	0.44
25	0.55	0.41	0.52	0.43	0.66	0.39	0.67	0.38	0.57	0.45	0.52	0.35
26	0.55	0.19	0.55	0.29	0.60	0.54	0.58	0.52	0.54	0.46	0.52	0.44
27	0.38	0.27	0.38	0.33	0.43	0.30	0.40	0.36	0.60	0.41	0.62	0.37
28	0.66	0.51	0.53	0.49	0.71	0.41	0.76	0.43	0.55	0.35	0.54	0.39
29	0.43	0.34	0.46	0.44	0.49	0.42	0.43	0.43	0.39	0.31	0.38	0.36
30	0.47	0.37	0.48	0.34	0.60	0.48	0.54	0.50	0.74	0.45	0.73	0.47
31	0.60	0.47	0.58	0.48	0.54	0.32	0.47	0.40	0.43	0.34	0.51	0.35
32	0.56	0.46	0.53	0.42	0.28	0.28	0.24	0.31	0.64	0.49	0.63	0.53
33	0.44	0.37	0.43	0.35	0.33	0.32	0.31	0.32	0.48	0.31	0.46	0.31
34	0.44	0.31	0.40	0.29	0.35	0.33	0.38	0.31	0.72	0.44	0.73	0.49
35	0.70	0.50	0.67	0.50	0.51	0.45	0.52	0.50	0.55	0.39	0.56	0.39
36	0.51	0.36	0.49	0.33	0.45	0.43	0.42	0.44	0.64	0.45	0.63	0.40
37	0.36	0.29	0.32	0.33	0.49	0.40	0.43	0.38	0.61	0.52	0.61	0.52
38	0.48	0.30	0.41	0.27	0.36	0.35	0.35	0.26	0.47	0.33	0.46	0.29
39	0.58	0.48	0.52	0.46	0.48	0.36	0.43	0.43	0.70	0.45	0.69	0.49
40	0.45	0.17	0.48	0.22	0.78	0.45	0.72	0.54	0.40	0.34	0.35	0.37
41	0.46	0.30	0.42	0.31	0.27	0.13	0.27	0.11	0.41	0.26	0.42	0.26
42	0.63	0.43	0.53	0.43	0.61	0.26	0.62	0.29	0.46	0.35	0.45	0.38
43	0.73	0.36	0.72	0.39	0.71	0.38	0.68	0.44	0.35	0.16	0.31	0.14
44	0.78	0.41	0.73	0.38	0.44	0.29	0.43	0.31	0.49	0.34	0.49	0.31
45	0.58	0.41	0.52	0.44	0.52	0.25	0.52	0.30	0.36	0.23	0.35	0.22
46	0.57	0.38	0.51	0.40	0.42	0.22	0.45	0.28	0.49	0.27	0.49	0.24
47	0.52	0.45	0.47	0.50	0.84	0.43	0.82	0.44	0.41	0.36	0.39	0.38
48	0.72	0.38	0.65	0.40	0.52	0.39	0.53	0.42	0.63	0.48	0.61	0.48
49	0.23	0.08	0.23	0.09	0.39	0.28	0.38	0.14	0.79	0.41	0.75	0.38
50	0.48	0.37	0.51	0.34	0.53	0.29	0.52	0.30	0.59	0.36	0.55	0.35
51	0.71	0.40	0.68	0.43	0.53	0.34	0.51	0.35	0.66	0.37	0.62	0.40

Item-by-item $p$ -value and Point Biserial for RLA												
STS	7				8				9			
Year	2011		2014		2012		2014		2011		2014	
Items	$p$ -value	Pt-Bis										
52	0.79	0.43	0.72	0.44	0.85	0.39	0.82	0.45	0.48	0.31	0.49	0.28
53	0.67	0.33	0.64	0.32	0.53	0.34	0.54	0.38	0.71	0.41	0.73	0.44
54	0.65	0.48	0.56	0.46	0.39	0.37	0.32	0.37	0.35	0.28	0.35	0.36
55	0.39	0.39	0.33	0.34	0.51	0.46	0.46	0.49	0.61	0.31	0.61	0.31
56	0.73	0.41	0.66	0.51	0.78	0.44	0.71	0.49	0.48	0.41	0.45	0.40
57	0.66	0.55	0.65	0.48	0.76	0.38	0.70	0.37	0.62	0.44	0.62	0.47
58	0.65	0.44	0.57	0.48	0.63	0.38	0.62	0.45	0.51	0.33	0.50	0.38
59	0.75	0.44	0.72	0.47	0.62	0.51	0.58	0.55	0.68	0.33	0.68	0.32
60	0.42	0.40	0.38	0.40	0.68	0.43	0.62	0.46	0.26	0.26	0.25	0.27
61	0.49	0.48	0.42	0.52	0.62	0.50	0.53	0.53	0.49	0.47	0.48	0.51
62	0.44	0.28	0.40	0.33	0.46	0.35	0.44	0.42	0.49	0.37	0.47	0.37
63	0.53	0.43	0.54	0.43	0.61	0.40	0.60	0.36	0.24	0.08	0.22	0.05
64	0.66	0.45	0.64	0.43	0.46	0.33	0.47	0.37	0.77	0.41	0.78	0.42
65	0.65	0.51	0.58	0.44	0.54	0.39	0.53	0.32	0.72	0.53	0.71	0.52
66	0.79	0.33	0.74	0.43	0.69	0.46	0.62	0.45	0.67	0.39	0.66	0.38
67	0.60	0.48	0.60	0.44	0.56	0.36	0.55	0.37	0.46	0.40	0.51	0.39
68	0.36	0.29	0.36	0.27	0.38	0.37	0.37	0.38	0.35	0.23	0.33	0.18
69	0.43	0.28	0.43	0.29	0.46	0.40	0.42	0.41	0.65	0.45	0.62	0.44
70	0.51	0.39	0.47	0.41	0.43	0.30	0.38	0.34	0.64	0.51	0.59	0.52
71	0.60	0.42	0.53	0.37	0.49	0.40	0.49	0.37	0.31	0.36	0.31	0.33
72	0.61	0.40	0.57	0.35	0.43	0.29	0.44	0.27	0.73	0.46	0.71	0.51
73	0.32	0.24	0.32	0.14	0.41	0.30	0.41	0.30	0.46	0.41	0.47	0.49
74	0.22	0.02	0.23	0.06	0.52	0.38	0.49	0.34	0.34	0.29	0.32	0.36
75	0.39	0.28	0.38	0.29	0.80	0.40	0.79	0.42	0.44	0.39	0.41	0.38

**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**

Item-by-item $p$ -value and Point Biserial for RLA									
STS	10				11				
Year	2012		2014		2011		2014		
Items	$p$ -value	Pt-Bis	Pt-Bis						
1	0.58	0.34	0.56	0.34	0.42	0.26	0.42	0.47	0.47
2	0.79	0.28	0.74	0.24	0.67	0.36	0.68	0.43	0.43
3	0.54	0.33	0.48	0.43	0.69	0.34	0.69	0.41	0.41
4	0.79	0.27	0.73	0.34	0.54	0.11	0.51	0.16	0.16
5	0.62	0.39	0.56	0.36	0.47	0.19	0.48	0.17	0.17
6	0.62	0.35	0.59	0.35	0.54	0.16	0.50	0.12	0.12
7	0.46	0.23	0.37	0.25	0.42	0.24	0.45	0.21	0.21
8	0.48	0.36	0.43	0.51	0.70	0.32	0.69	0.19	0.19
9	0.73	0.33	0.68	0.33	0.41	0.30	0.40	0.29	0.29
10	0.71	0.48	0.67	0.49	0.73	0.40	0.68	0.53	0.53
11	0.54	0.43	0.54	0.35	0.64	0.43	0.64	0.49	0.49
12	0.56	0.34	0.53	0.41	0.80	0.22	0.72	0.24	0.24
13	0.82	0.43	0.77	0.51	0.68	0.32	0.67	0.32	0.32
14	0.73	0.35	0.70	0.40	0.54	0.09	0.50	0.12	0.12
15	0.49	0.21	0.47	0.30	0.33	0.33	0.44	0.33	0.33
16	0.84	0.45	0.81	0.53	0.45	0.28	0.44	0.32	0.32
17	0.35	0.40	0.31	0.31	0.83	0.45	0.79	0.51	0.51
18	0.43	0.25	0.47	0.26	0.39	0.37	0.42	0.44	0.44
19	0.77	0.47	0.73	0.54	0.59	0.27	0.60	0.42	0.42
20	0.69	0.35	0.65	0.40	0.56	0.43	0.60	0.48	0.48
21	0.66	0.45	0.60	0.51	0.55	0.25	0.56	0.32	0.32

Item-by-item $p$ -value and Point Biserial for RLA								
STS	10				11			
Year	2012		2014		2011		2014	
Items	$p$ -value	Pt-Bis						
22	0.32	0.17	0.32	0.29	0.44	0.30	0.44	0.32
23	0.72	0.44	0.72	0.44	0.44	0.26	0.45	0.39
24	0.65	0.36	0.60	0.42	0.49	0.10	0.50	0.08
25	0.72	0.43	0.65	0.43	0.46	0.26	0.38	0.14
26	0.53	0.33	0.47	0.30	0.53	0.34	0.52	0.45
27	0.90	0.37	0.86	0.45	0.43	0.20	0.44	0.25
28	0.88	0.41	0.84	0.41	0.50	0.17	0.51	0.19
29	0.43	0.29	0.39	0.26	0.60	0.45	0.60	0.47
30	0.36	0.27	0.31	0.15	0.31	0.13	0.27	0.21
31	0.55	0.31	0.58	0.31	0.39	0.17	0.42	0.17
32	0.44	0.21	0.43	0.24	0.33	0.24	0.32	0.06
33	0.70	0.38	0.61	0.45	0.59	0.38	0.58	0.49
34	0.58	0.28	0.50	0.38	0.47	0.41	0.49	0.45
35	0.60	0.40	0.54	0.38	0.49	0.10	0.49	0.11
36	0.72	0.49	0.67	0.52	0.77	0.44	0.77	0.43
37	0.63	0.40	0.58	0.46	0.52	0.39	0.58	0.27
38	0.40	0.32	0.35	0.38	0.41	0.39	0.46	0.38
39	0.82	0.45	0.74	0.58	0.75	0.36	0.75	0.19
40	0.66	0.44	0.55	0.45	0.70	0.38	0.70	0.39
41	0.38	0.27	0.35	0.29	0.79	0.49	0.77	0.56
42	0.60	0.03	0.57	0.16	0.64	0.25	0.61	0.24
43	0.71	0.41	0.63	0.41	0.70	0.31	0.57	0.41
44	0.62	0.27	0.53	0.30	0.90	0.37	0.87	0.49
45	0.64	0.43	0.59	0.45	0.48	0.36	0.50	0.37
46	0.54	0.34	0.48	0.46	0.65	0.33	0.54	0.47
47	0.51	0.38	0.46	0.47	0.60	0.40	0.63	0.42
48	0.68	0.44	0.59	0.45	0.53	0.35	0.60	0.36
49	0.44	0.32	0.37	0.31	0.68	0.46	0.62	0.50
50	0.54	0.30	0.52	0.26	0.58	0.44	0.63	0.43
51	0.40	0.23	0.31	0.18	0.62	0.39	0.55	0.34
52	0.55	0.27	0.50	0.28	0.41	0.25	0.49	0.31
53	0.42	0.25	0.38	0.29	0.55	0.39	0.56	0.40
54	0.55	0.46	0.48	0.34	0.45	0.37	0.43	0.21
55	0.35	0.24	0.31	0.27	0.60	0.42	0.56	0.35
56	0.50	0.37	0.47	0.33	0.53	0.39	0.48	0.55
57	0.40	0.41	0.37	0.42	0.23	0.27	0.27	0.49
58	0.58	0.40	0.51	0.37	0.69	0.43	0.62	0.53
59	0.79	0.35	0.70	0.41	0.46	0.46	0.48	0.44
60	0.71	0.33	0.60	0.45	0.70	0.45	0.78	0.48
61	0.34	0.25	0.34	0.31	0.64	0.43	0.63	0.48
62	0.34	0.22	0.29	0.16	0.61	0.44	0.58	0.57
63	0.88	0.40	0.82	0.50	0.38	0.28	0.37	0.26
64	0.75	0.43	0.67	0.43	0.45	0.19	0.42	0.24
65	0.59	0.28	0.61	0.30	0.74	0.43	0.74	0.57
66	0.28	0.23	0.23	0.19	0.72	0.41	0.73	0.48
67	0.61	0.36	0.59	0.39	0.52	0.34	0.49	0.31
68	0.66	0.47	0.56	0.38	0.53	0.44	0.53	0.40
69	0.38	0.29	0.37	0.26	0.54	0.33	0.55	0.45
70	0.65	0.36	0.54	0.33	0.39	0.12	0.40	0.19
71	0.55	0.34	0.48	0.45	0.73	0.41	0.68	0.48
72	0.56	0.30	0.49	0.41	0.26	0.04	0.34	0.08
73	0.34	0.32	0.28	0.38	0.60	0.43	0.67	0.35
74	0.70	0.36	0.64	0.34	0.62	0.41	0.60	0.41
75	0.35	0.16	0.29	0.16	0.14	-0.05	0.19	0.13

## 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		
<b>Grade 2</b>								
1. Word Analysis and Vocabulary Development	22	1.00	.	.	.	.	0.83	1.94
2. Reading Comprehension	15	0.75	1.00	.	.	.	0.78	1.68
3. Literary Response and Analysis	6	0.65	0.64	1.00	.	.	0.59	1.06
4. Written Conventions	14	0.77	0.73	0.61	1.00	.	0.80	1.61
5. Writing Strategies	8	0.59	0.61	0.52	0.63	1.00	0.56	1.30
<b>Grade 3</b>								
1. Word Analysis and Vocabulary Development	20	1.00	.	.	.	.	0.76	1.97
2. Reading Comprehension	15	0.66	1.00	.	.	.	0.63	1.78
3. Literary Response and Analysis	8	0.60	0.55	1.00	.	.	0.45	1.26
4. Written Conventions	13	0.63	0.58	0.49	1.00	.	0.68	1.65
5. Writing Strategies	9	0.58	0.54	0.51	0.58	1.00	0.57	1.37
<b>Grade 4</b>								
1. Word Analysis and Vocabulary Development	18	1.00	.	.	.	.	0.79	1.83
2. Reading Comprehension	15	0.70	1.00	.	.	.	0.74	1.75
3. Literary Response and Analysis	9	0.55	0.61	1.00	.	.	0.53	1.36
4. Written Conventions	18	0.74	0.66	0.55	1.00	.	0.82	1.82
5. Writing Strategies	15	0.69	0.65	0.51	0.71	1.00	0.68	1.75
<b>Grade 5</b>								
1. Word Analysis and Vocabulary Development	14	1.00	.	.	.	.	0.66	1.70
2. Reading Comprehension	16	0.68	1.00	.	.	.	0.69	1.84
3. Literary Response and Analysis	12	0.66	0.67	1.00	.	.	0.72	1.55
4. Written Conventions	17	0.62	0.63	0.61	1.00	.	0.75	1.85
5. Writing Strategies	16	0.63	0.66	0.63	0.69	1.00	0.71	1.82
<b>Grade 6</b>								
1. Word Analysis and Vocabulary Development	13	1.00	.	.	.	.	0.68	1.61
2. Reading Comprehension	17	0.67	1.00	.	.	.	0.71	1.90
3. Literary Response and Analysis	12	0.62	0.62	1.00	.	.	0.63	1.59
4. Written Conventions	16	0.68	0.65	0.64	1.00	.	0.79	1.76
5. Writing Strategies	17	0.66	0.65	0.63	0.69	1.00	0.69	1.89
<b>Grade 7</b>								
1. Word Analysis and Vocabulary Development	11	1.00	.	.	.	.	0.65	1.47
2. Reading Comprehension	18	0.69	1.00	.	.	.	0.71	1.93
3. Literary Response and Analysis	13	0.65	0.72	1.00	.	.	0.68	1.58
4. Written Conventions	16	0.62	0.64	0.58	1.00	.	0.74	1.76
5. Writing Strategies	17	0.65	0.69	0.64	0.72	1.00	0.71	1.85
<b>Grade 8</b>								
1. Word Analysis and Vocabulary Development	9	1.00	.	.	.	.	0.71	1.29
2. Reading Comprehension	18	0.64	1.00	.	.	.	0.74	1.91
3. Literary Response and Analysis	15	0.64	0.61	1.00	.	.	0.69	1.78
4. Written Conventions	16	0.65	0.60	0.57	1.00	.	0.76	1.73
5. Writing Strategies	17	0.65	0.61	0.63	0.71	1.00	0.67	1.89

<b>Grade 9</b>		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>		
1. Word Analysis and Vocabulary Development	8	1.00	.	.	.	.	0.66	1.10
2. Reading Comprehension	18	0.67	1.00	.	.	.	0.75	1.85
3. Literary Response and Analysis	16	0.60	0.68	1.00	.	.	0.66	1.77
4. Written Conventions	13	0.57	0.59	0.52	1.00	.	0.65	1.60
5. Writing Strategies	20	0.62	0.65	0.59	0.66	1.00	0.75	2.02
<b>Grade 10</b>		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>		
1. Word Analysis and Vocabulary Development	8	1.00	.	.	.	.	0.66	1.12
2. Reading Comprehension	18	0.70	1.00	.	.	.	0.78	1.88
3. Literary Response and Analysis	16	0.68	0.71	1.00	.	.	0.68	1.77
4. Written Conventions	13	0.63	0.64	0.59	1.00	.	0.65	1.64
5. Writing Strategies	20	0.62	0.60	0.56	0.69	1.00	0.68	2.06
<b>Grade 11</b>		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>		
1. Word Analysis and Vocabulary Development	8	1.00	.	.	.	.	0.64	1.19
2. Reading Comprehension	19	0.63	1.00	.	.	.	0.58	2.05
3. Literary Response and Analysis	17	0.70	0.58	1.00	.	.	0.65	1.90
4. Written Conventions	9	0.63	0.47	0.57	1.00	.	0.57	1.31
5. Writing Strategies	22	0.72	0.59	0.70	0.65	1.00	0.79	2.10

**Table 8.B.2 Reliabilities and SEMs by Gender**

<b>Content Area</b>	<b>STS *</b>	<b>Male</b>			<b>Female</b>		
		<b>N</b>	<b>Reliab.</b>	<b>SEM</b>	<b>N</b>	<b>Reliab.</b>	<b>SEM</b>
<b>Reading/ Language Arts</b>	2	979	0.93	3.54	947	0.93	3.44
	3	695	0.89	3.67	691	0.89	3.64
	4	419	0.93	3.89	383	0.92	3.82
	5	348	0.91	3.96	293	0.92	3.93
	6	243	0.92	3.96	236	0.91	3.93
	7	229	0.92	3.90	225	0.92	3.85
	8	243	0.92	3.91	194	0.91	3.90
	9	581	0.91	3.85	372	0.91	3.77
	10	226	0.91	3.89	162	0.91	3.83
	11	99	0.91	3.94	83	0.88	3.91

\* STS named by number only are grade-level tests.

**Table 8.B.3 Reliabilities and SEMs by Economic Status**

<b>Content Area</b>	<b>STS *</b>	<b>Not Disadvantaged</b>			<b>Disadvantaged</b>		
		<b>N</b>	<b>Reliab.</b>	<b>SEM</b>	<b>N</b>	<b>Reliab.</b>	<b>SEM</b>
<b>Reading/ Language Arts</b>	2	87	0.95	3.46	1,720	0.93	3.49
	3	72	0.91	3.54	1,242	0.89	3.66
	4	57	0.94	3.81	734	0.93	3.86
	5	45	0.91	3.97	595	0.92	3.95
	6	41	0.93	3.90	429	0.92	3.95
	7	36	0.94	3.81	413	0.92	3.89
	8	43	0.90	3.93	385	0.92	3.90
	9	163	0.90	3.84	776	0.91	3.82
	10	58	0.91	3.87	310	0.91	3.87
	11	28	0.86	3.86	148	0.90	3.94

\* STS named by number only are grade-level tests.

**Table 8.B.4 Reliabilities and SEMs by Special Services**

Content Area	STS *	No Special Services			Special Services		
		N	Reliab.	SEM	N	Reliab.	SEM
Reading/Language Arts	2	1,852	0.93	3.49	78	0.88	3.68
	3	1,337	0.89	3.66	51	0.83	3.68
	4	778	0.93	3.86	25	0.93	3.85
	5	627	0.92	3.95	18	0.82	3.99
	6	474	0.92	3.95	8	–	–
	7	455	0.92	3.88	5	–	–
	8	433	0.92	3.91	6	–	–
	9	955	0.91	3.82	1	–	–
	10	387	0.91	3.87	1	–	–
	11	183	0.90	3.93	0	–	–

\* STS named by number only are grade-level tests.

**Table 8.B.5 Reliabilities and SEMs by Attendance in U.S. Schools**

Content Area	STS *	In U.S. Schools < 12 Months			In U.S. Schools ≥ 12 Months		
		N	Reliab.	SEM	N	Reliab.	SEM
Reading/Language Arts	2	526	0.94	3.54	1,404	0.92	3.47
	3	466	0.90	3.64	922	0.89	3.66
	4	470	0.92	3.89	333	0.93	3.79
	5	450	0.92	3.93	195	0.91	3.96
	6	365	0.92	3.92	117	0.89	3.97
	7	390	0.92	3.85	70	0.88	3.99
	8	416	0.92	3.90	23	0.86	3.99
	9	916	0.91	3.82	40	0.93	3.86
	10	338	0.91	3.86	50	0.90	3.93
	11	161	0.89	3.91	22	0.85	3.99

\* STS named by number only are grade-level tests.

**Table 8.B.6 Reliabilities and SEMs by EL Program Participation**

Content Area	STS *	Primary language instruction and ELD and/or SDAIE instruction			ELD instruction only			SDAIE instruction only			ELD instr. and SDAIE instr. but not primary language instr.			Other EL Instruct. Services			None (EL only)		
		N	Rel.	SEM	N	Rel.	SEM	N	Rel.	SEM	N	Rel.	SEM	N	Rel.	SEM	N	Rel.	SEM
Reading/ Language Arts	2	1,642	0.93	3.46	12	0.88	3.64	140	0.87	3.65	117	0.92	3.62	2	–	–	8	–	–
	3	1,117	0.89	3.65	11	0.91	3.63	117	0.87	3.67	123	0.89	3.65	1	–	–	13	0.86	3.68
	4	491	0.93	3.80	8	–	–	102	0.92	3.93	178	0.92	3.90	1	–	–	10	–	–
	5	294	0.91	3.95	18	0.93	3.87	113	0.92	3.93	191	0.92	3.93	0	–	–	12	0.93	3.87
	6	161	0.90	3.97	5	–	–	19	0.89	3.93	282	0.93	3.91	1	–	–	11	0.90	3.95
	7	137	0.91	3.92	11	0.92	3.93	3	–	–	287	0.92	3.86	2	–	–	9	–	–
	8	81	0.91	3.90	10	–	–	4	–	–	318	0.92	3.91	0	–	–	13	0.93	3.88
	9	246	0.92	3.85	6	–	–	8	–	–	621	0.91	3.81	10	–	–	22	0.87	3.75
	10	151	0.90	3.92	8	–	–	7	–	–	198	0.91	3.83	1	–	–	15	0.92	3.81
	11	56	0.89	4.01	6	–	–	5	–	–	103	0.90	3.89	1	–	–	5	–	–

\* STS named by number only are grade-level tests.

**Table 8.B.7 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
<b>Grade 2</b>									
1. Word Analysis and Vocabulary Development	22	0.83	1.97	0.83	1.91	0.86	1.93	0.83	1.94
2. Reading Comprehension	15	0.78	1.70	0.78	1.66	0.80	1.70	0.78	1.68
3. Literary Response and Analysis	6	0.56	1.08	0.61	1.03	0.69	1.02	0.59	1.06
4. Written Conventions	14	0.79	1.64	0.81	1.58	0.83	1.59	0.80	1.61
5. Writing Strategies	8	0.56	1.30	0.56	1.31	0.66	1.25	0.55	1.31
<b>Grade 3</b>									
1. Word Analysis and Vocabulary Development	20	0.76	1.98	0.76	1.96	0.77	1.87	0.76	1.98
2. Reading Comprehension	15	0.59	1.79	0.66	1.77	0.66	1.75	0.63	1.78
3. Literary Response and Analysis	8	0.44	1.26	0.45	1.25	0.55	1.20	0.44	1.26
4. Written Conventions	13	0.66	1.66	0.69	1.64	0.73	1.60	0.68	1.65
5. Writing Strategies	9	0.59	1.37	0.55	1.38	0.58	1.36	0.58	1.37
<b>Grade 4</b>									
1. Word Analysis and Vocabulary Development	18	0.79	1.86	0.77	1.81	0.79	1.84	0.79	1.83
2. Reading Comprehension	15	0.74	1.76	0.72	1.74	0.80	1.68	0.74	1.75
3. Literary Response and Analysis	9	0.50	1.37	0.53	1.34	0.60	1.34	0.53	1.36
4. Written Conventions	18	0.82	1.85	0.80	1.79	0.86	1.79	0.81	1.83
5. Writing Strategies	15	0.70	1.73	0.64	1.76	0.67	1.76	0.69	1.74
<b>Grade 5</b>									
1. Word Analysis and Vocabulary Development	14	0.65	1.71	0.66	1.70	0.63	1.71	0.67	1.70
2. Reading Comprehension	16	0.67	1.84	0.69	1.83	0.75	1.81	0.68	1.84
3. Literary Response and Analysis	12	0.72	1.55	0.72	1.55	0.74	1.54	0.72	1.55
4. Written Conventions	17	0.73	1.87	0.75	1.83	0.68	1.87	0.75	1.85
5. Writing Strategies	16	0.70	1.81	0.70	1.82	0.57	1.85	0.72	1.81
<b>Grade 6</b>									
1. Word Analysis and Vocabulary Development	13	0.66	1.62	0.69	1.61	0.63	1.61	0.69	1.61
2. Reading Comprehension	17	0.70	1.91	0.71	1.89	0.70	1.90	0.71	1.89
3. Literary Response and Analysis	12	0.64	1.60	0.60	1.58	0.71	1.55	0.62	1.59
4. Written Conventions	16	0.78	1.78	0.78	1.74	0.84	1.72	0.78	1.77
5. Writing Strategies	17	0.70	1.88	0.67	1.90	0.72	1.86	0.69	1.90
<b>Grade 7</b>									
1. Word Analysis and Vocabulary Development	11	0.67	1.46	0.63	1.48	0.66	1.48	0.65	1.47
2. Reading Comprehension	18	0.72	1.93	0.71	1.94	0.75	1.91	0.71	1.94
3. Literary Response and Analysis	13	0.68	1.60	0.69	1.57	0.71	1.57	0.69	1.58
4. Written Conventions	16	0.73	1.79	0.75	1.73	0.76	1.73	0.74	1.76
5. Writing Strategies	17	0.68	1.87	0.74	1.82	0.80	1.80	0.70	1.85
<b>Grade 8</b>									
1. Word Analysis and Vocabulary Development	9	0.72	1.28	0.70	1.29	0.72	1.28	0.71	1.28
2. Reading Comprehension	18	0.74	1.92	0.76	1.90	0.69	1.92	0.75	1.90
3. Literary Response and Analysis	15	0.72	1.76	0.63	1.80	0.68	1.75	0.70	1.77
4. Written Conventions	16	0.76	1.75	0.73	1.69	0.77	1.74	0.76	1.73
5. Writing Strategies	17	0.67	1.88	0.65	1.90	0.62	1.91	0.68	1.89
<b>Grade 9</b>									
1. Word Analysis and Vocabulary Development	8	0.69	1.12	0.63	1.06	0.62	1.10	0.67	1.10
2. Reading Comprehension	18	0.76	1.86	0.73	1.83	0.69	1.87	0.76	1.84
3. Literary Response and Analysis	16	0.66	1.79	0.65	1.73	0.65	1.76	0.66	1.77

Subscore Area	No. of Items	Male		Female		Not Econ. Dis.		Econ. Dis.	
		Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
4. Written Conventions	13	0.62	1.62	0.66	1.57	0.62	1.61	0.65	1.60
5. Writing Strategies	20	0.73	2.02	0.76	2.01	0.71	2.04	0.75	2.01
<b>Grade 10</b>									
1. Word Analysis and Vocabulary Development	8	0.67	1.15	0.62	1.08	0.74	1.10	0.64	1.12
2. Reading Comprehension	18	0.76	1.90	0.79	1.85	0.80	1.88	0.77	1.87
3. Literary Response and Analysis	16	0.66	1.79	0.67	1.73	0.63	1.77	0.67	1.77
4. Written Conventions	13	0.65	1.65	0.64	1.63	0.62	1.67	0.66	1.63
5. Writing Strategies	20	0.69	2.04	0.65	2.08	0.65	2.03	0.68	2.06
<b>Grade 11</b>									
1. Word Analysis and Vocabulary Development	8	0.68	1.21	0.57	1.14	0.58	1.13	0.66	1.19
2. Reading Comprehension	19	0.60	2.04	0.54	2.05	0.13	2.11	0.61	2.04
3. Literary Response and Analysis	17	0.66	1.92	0.59	1.87	0.58	1.87	0.65	1.90
4. Written Conventions	9	0.61	1.30	0.48	1.32	0.72	1.19	0.53	1.33
5. Writing Strategies	22	0.78	2.13	0.79	2.08	0.75	2.01	0.78	2.12

**Table 8.B.8 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
<b>Grade 2</b>									
1. Word Analysis and Vocabulary Development	22	0.83	1.93	0.74	2.12	0.84	2.03	0.80	1.90
2. Reading Comprehension	15	0.78	1.68	0.71	1.76	0.79	1.70	0.77	1.68
3. Literary Response and Analysis	6	0.60	1.05	0.30	1.15	0.59	1.08	0.57	1.05
4. Written Conventions	14	0.80	1.61	0.63	1.72	0.83	1.60	0.76	1.61
5. Writing Strategies	8	0.56	1.30	0.42	1.25	0.56	1.28	0.55	1.31
<b>Grade 3</b>									
1. Word Analysis and Vocabulary Development	20	0.75	1.97	0.71	2.04	0.78	1.97	0.74	1.97
2. Reading Comprehension	15	0.63	1.78	0.58	1.74	0.65	1.77	0.62	1.78
3. Literary Response and Analysis	8	0.45	1.25	-0.02	1.31	0.49	1.24	0.42	1.26
4. Written Conventions	13	0.68	1.65	0.50	1.63	0.67	1.63	0.67	1.65
5. Writing Strategies	9	0.57	1.38	0.50	1.35	0.54	1.38	0.59	1.37
<b>Grade 4</b>									
1. Word Analysis and Vocabulary Development	18	0.79	1.83	0.81	1.86	0.78	1.86	0.79	1.78
2. Reading Comprehension	15	0.74	1.75	0.77	1.72	0.70	1.78	0.79	1.71
3. Literary Response and Analysis	9	0.53	1.36	0.21	1.41	0.52	1.36	0.54	1.35
4. Written Conventions	18	0.82	1.82	0.81	1.88	0.83	1.83	0.80	1.80
5. Writing Strategies	15	0.68	1.75	0.76	1.62	0.67	1.75	0.68	1.73
<b>Grade 5</b>									
1. Word Analysis and Vocabulary Development	14	0.67	1.70	0.36	1.72	0.67	1.70	0.64	1.71
2. Reading Comprehension	16	0.69	1.83	0.50	1.86	0.69	1.83	0.70	1.82
3. Literary Response and Analysis	12	0.72	1.56	0.55	1.48	0.73	1.55	0.70	1.56
4. Written Conventions	17	0.75	1.85	0.66	1.90	0.76	1.84	0.73	1.85
5. Writing Strategies	16	0.71	1.81	0.61	1.85	0.71	1.80	0.70	1.83
<b>Grade 6</b>									
1. Word Analysis and Vocabulary Development	13	0.68	1.61	–	–	0.69	1.61	0.66	1.61
2. Reading Comprehension	17	0.70	1.90	–	–	0.73	1.89	0.62	1.90
3. Literary Response and Analysis	12	0.63	1.59	–	–	0.65	1.58	0.59	1.60
4. Written Conventions	16	0.79	1.76	–	–	0.80	1.75	0.73	1.77
5. Writing Strategies	17	0.70	1.89	–	–	0.71	1.87	0.63	1.92
<b>Grade 7</b>									
1. Word Analysis and Vocabulary Development	11	0.64	1.47	–	–	0.65	1.46	0.58	1.53
2. Reading Comprehension	18	0.71	1.94	–	–	0.72	1.93	0.64	1.97
3. Literary Response and Analysis	13	0.68	1.58	–	–	0.69	1.57	0.60	1.65
4. Written Conventions	16	0.74	1.76	–	–	0.75	1.75	0.68	1.79
5. Writing Strategies	17	0.71	1.85	–	–	0.73	1.83	0.53	1.93
<b>Grade 8</b>									
1. Word Analysis and Vocabulary Development	9	0.71	1.29	–	–	0.72	1.28	0.45	1.41
2. Reading Comprehension	18	0.74	1.91	–	–	0.75	1.91	0.06	1.95
3. Literary Response and Analysis	15	0.69	1.78	–	–	0.69	1.77	0.64	1.74
4. Written Conventions	16	0.76	1.73	–	–	0.76	1.72	0.62	1.88
5. Writing Strategies	17	0.67	1.89	–	–	0.66	1.90	0.71	1.86

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
<b>Grade 9</b>									
1. Word Analysis and Vocabulary Development	8	0.67	1.10	–	–	0.66	1.10	0.74	1.14
2. Reading Comprehension	18	0.75	1.85	–	–	0.75	1.85	0.80	1.85
3. Literary Response and Analysis	16	0.66	1.77	–	–	0.66	1.77	0.73	1.82
4. Written Conventions	13	0.65	1.60	–	–	0.64	1.60	0.74	1.58
5. Writing Strategies	20	0.75	2.02	–	–	0.75	2.02	0.76	2.02
<b>Grade 10</b>									
1. Word Analysis and Vocabulary Development	8	0.65	1.12	–	–	0.66	1.10	0.63	1.21
2. Reading Comprehension	18	0.78	1.88	–	–	0.78	1.87	0.68	1.94
3. Literary Response and Analysis	16	0.67	1.77	–	–	0.67	1.77	0.68	1.77
4. Written Conventions	13	0.65	1.64	–	–	0.64	1.64	0.67	1.63
5. Writing Strategies	20	0.68	2.06	–	–	0.68	2.06	0.64	2.05
<b>Grade 11</b>									
1. Word Analysis and Vocabulary Development	8	0.64	1.19	–	–	0.61	1.17	0.54	1.31
2. Reading Comprehension	19	0.58	2.05	–	–	0.58	2.04	0.31	2.04
3. Literary Response and Analysis	17	0.65	1.90	–	–	0.59	1.90	0.62	1.88
4. Written Conventions	9	0.57	1.31	–	–	0.55	1.29	0.51	1.38
5. Writing Strategies	22	0.79	2.10	–	–	0.76	2.10	0.71	2.10

Table 8.B.9 Subscore Reliabilities and SEM for RLA by EL Program Participation

Subscore Reliabilities and SEM for RLA by EL Program Participation													
Subscore Area	No. of Items	Primary language instruction and ELD and/or SDAIE instruction		ELD instruction only		SDAIE instruction only		ELD instr. and SDAIE instr. but not primary language instr.		Other EL Instruct. Services		None (EL only)	
		Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
<b>Grade 2</b>													
1. Word Analysis and Vocabulary Development	22	0.81	1.90	0.62	2.23	0.74	2.13	0.77	2.12	–	–	–	–
2. Reading Comprehension	15	0.78	1.67	0.69	1.74	0.61	1.77	0.73	1.74	–	–	–	–
3. Literary Response and Analysis	6	0.58	1.05	0.49	1.08	0.41	1.10	0.47	1.13	–	–	–	–
4. Written Conventions	14	0.77	1.61	0.72	1.59	0.64	1.66	0.78	1.62	–	–	–	–
5. Writing Strategies	8	0.56	1.31	0.24	1.26	0.36	1.24	0.38	1.29	–	–	–	–
<b>Grade 3</b>													
1. Word Analysis and Vocabulary Development	20	0.75	1.97	0.79	2.04	0.75	2.00	0.78	1.98	–	–	0.67	2.09
2. Reading Comprehension	15	0.64	1.77	0.49	1.75	0.56	1.78	0.61	1.77	–	–	0.46	1.77
3. Literary Response and Analysis	8	0.44	1.26	0.56	1.22	0.42	1.26	0.49	1.24	–	–	0.34	1.18
4. Written Conventions	13	0.67	1.65	0.82	1.51	0.58	1.65	0.63	1.62	–	–	0.61	1.67
5. Writing Strategies	9	0.58	1.37	0.62	1.35	0.55	1.35	0.45	1.41	–	–	0.58	1.37
<b>Grade 4</b>													
1. Word Analysis and Vocabulary Development	18	0.78	1.79	–	–	0.78	1.90	0.77	1.88	–	–	–	–
2. Reading Comprehension	15	0.76	1.72	–	–	0.71	1.79	0.70	1.77	–	–	–	–
3. Literary Response and Analysis	9	0.53	1.35	–	–	0.55	1.36	0.50	1.36	–	–	–	–
4. Written Conventions	18	0.81	1.80	–	–	0.80	1.89	0.83	1.83	–	–	–	–
5. Writing Strategies	15	0.69	1.73	–	–	0.65	1.75	0.65	1.77	–	–	–	–
<b>Grade 5</b>													
1. Word Analysis and Vocabulary Development	14	0.64	1.71	0.68	1.72	0.67	1.71	0.68	1.69	–	–	0.83	1.56
2. Reading Comprehension	16	0.68	1.84	0.71	1.74	0.68	1.82	0.69	1.84	–	–	0.77	1.74
3. Literary Response and Analysis	12	0.72	1.55	0.70	1.53	0.75	1.54	0.71	1.57	–	–	0.69	1.54
4. Written Conventions	17	0.72	1.85	0.78	1.86	0.77	1.85	0.75	1.85	–	–	0.83	1.81
5. Writing Strategies	16	0.71	1.82	0.78	1.79	0.68	1.82	0.72	1.79	–	–	0.56	1.82

Subscore Reliabilities and SEM for RLA by EL Program Participation													
Subscore Area	No. of Items	Primary language instruction and ELD and/or SDAIE instruction		ELD instruction only		SDAIE instruction only		ELD instr. and SDAIE instr. but not primary language instr.		Other EL Instruct. Services		None (EL only)	
		Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
<b>Grade 6</b>													
1. Word Analysis and Vocabulary Development	13	0.65	1.62	–	–	0.76	1.55	0.69	1.61	–	–	0.59	1.66
2. Reading Comprehension	17	0.64	1.90	–	–	0.47	1.91	0.72	1.89	–	–	0.72	1.87
3. Literary Response and Analysis	12	0.59	1.59	–	–	0.51	1.55	0.67	1.57	–	–	0.45	1.64
4. Written Conventions	16	0.75	1.77	–	–	0.79	1.73	0.80	1.74	–	–	0.58	1.82
5. Writing Strategies	17	0.66	1.92	–	–	0.60	1.94	0.71	1.87	–	–	0.80	1.79
<b>Grade 7</b>													
1. Word Analysis and Vocabulary Development	11	0.62	1.50	0.83	1.34	–	–	0.63	1.47	–	–	–	–
2. Reading Comprehension	18	0.73	1.93	0.60	2.01	–	–	0.71	1.93	–	–	–	–
3. Literary Response and Analysis	13	0.67	1.60	0.44	1.71	–	–	0.69	1.57	–	–	–	–
4. Written Conventions	16	0.72	1.77	0.82	1.75	–	–	0.75	1.75	–	–	–	–
5. Writing Strategies	17	0.64	1.89	0.57	1.85	–	–	0.74	1.83	–	–	–	–
<b>Grade 8</b>													
1. Word Analysis and Vocabulary Development	9	0.67	1.29	–	–	–	–	0.71	1.29	–	–	0.77	1.24
2. Reading Comprehension	18	0.67	1.92	–	–	–	–	0.74	1.91	–	–	0.84	1.83
3. Literary Response and Analysis	15	0.65	1.78	–	–	–	–	0.69	1.78	–	–	0.81	1.71
4. Written Conventions	16	0.73	1.73	–	–	–	–	0.76	1.73	–	–	0.67	1.83
5. Writing Strategies	17	0.72	1.87	–	–	–	–	0.65	1.90	–	–	0.76	1.79
<b>Grade 9</b>													
1. Word Analysis and Vocabulary Development	8	0.70	1.11	–	–	–	–	0.65	1.10	–	–	0.73	0.90
2. Reading Comprehension	18	0.75	1.87	–	–	–	–	0.75	1.84	–	–	0.75	1.77
3. Literary Response and Analysis	16	0.67	1.79	–	–	–	–	0.65	1.77	–	–	0.69	1.71
4. Written Conventions	13	0.61	1.63	–	–	–	–	0.66	1.59	–	–	0.57	1.62
5. Writing Strategies	20	0.76	2.02	–	–	–	–	0.75	2.02	–	–	0.65	2.02

<b>Subscore Reliabilities and SEM for RLA by EL Program Participation</b>													
<b>Subscore Area</b>	<b>No. of Items</b>	<b>Primary language instruction and ELD and/or SDAIE instruction</b>		<b>ELD instruction only</b>		<b>SDAIE instruction only</b>		<b>ELD instr. and SDAIE instr. but not primary language instr.</b>		<b>Other EL Instruct. Services</b>		<b>None (EL only)</b>	
		<b>Reliab.</b>	<b>SEM</b>	<b>Reliab.</b>	<b>SEM</b>	<b>Reliab.</b>	<b>SEM</b>	<b>Reliab.</b>	<b>SEM</b>	<b>Reliab.</b>	<b>SEM</b>	<b>Reliab.</b>	<b>SEM</b>
<b>Grade 10</b>													
1. Word Analysis and Vocabulary Development	8	0.65	1.19	–	–	–	–	0.68	1.06	–	–	0.26	1.14
2. Reading Comprehension	18	0.74	1.91	–	–	–	–	0.79	1.84	–	–	0.74	1.92
3. Literary Response and Analysis	16	0.66	1.78	–	–	–	–	0.67	1.76	–	–	0.70	1.76
4. Written Conventions	13	0.65	1.64	–	–	–	–	0.66	1.63	–	–	0.64	1.63
5. Writing Strategies	20	0.67	2.05	–	–	–	–	0.68	2.06	–	–	0.79	1.95
<b>Grade 11</b>													
1. Word Analysis and Vocabulary Development	8	0.59	1.27	–	–	–	–	0.62	1.16	–	–	–	–
2. Reading Comprehension	19	0.36	2.09	–	–	–	–	0.61	2.03	–	–	–	–
3. Literary Response and Analysis	17	0.71	1.89	–	–	–	–	0.60	1.90	–	–	–	–
4. Written Conventions	9	0.43	1.38	–	–	–	–	0.60	1.24	–	–	–	–
5. Writing Strategies	22	0.81	2.12	–	–	–	–	0.77	2.09	–	–	–	–

**Table 8.B.10 Reliability of Classification for RLA, Grade Two**

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
<b>Decision Accuracy</b>	0–17	0.05	0.02	0.00	0.00	0.00	0.08
	18–33	0.01	0.25	0.03	0.00	0.00	0.30
	34–47	0.00	0.03	0.26	0.03	0.00	0.33
<b>All-forms Average</b>	48–55	0.00	0.00	0.04	0.13	0.02	0.20
	56–65	0.00	0.00	0.00	0.02	0.08	0.10
<b>Estimated Proportion Correctly Classified: Total = 0.78, Proficient &amp; Above = 0.93</b>							
<b>Decision Consistency</b>	0–17	0.05	0.03	0.00	0.00	0.00	0.08
	18–33	0.03	0.22	0.05	0.00	0.00	0.30
	34–47	0.00	0.05	0.23	0.05	0.00	0.33
<b>Alternate Form</b>	48–55	0.00	0.00	0.05	0.11	0.04	0.20
	56–65	0.00	0.00	0.00	0.03	0.08	0.10
<b>Estimated Proportion Consistently Classified: Total = 0.69, Proficient &amp; Above = 0.90</b>							

**Table 8.B.11 Reliability of Classification for RLA, Grade Three**

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
<b>Decision Accuracy</b>	0–17	0.03	0.04	0.00	0.00	0.00	0.07
	18–29	0.02	0.29	0.05	0.00	0.00	0.36
	30–43	0.00	0.05	0.30	0.03	0.00	0.38
<b>All-forms Average</b>	44–52	0.00	0.00	0.04	0.10	0.01	0.15
	53–65	0.00	0.00	0.00	0.02	0.03	0.05
<b>Estimated Proportion Correctly Classified: Total = 0.75, Proficient &amp; Above = 0.94</b>							
<b>Decision Consistency</b>	0–17	0.03	0.04	0.00	0.00	0.00	0.07
	18–29	0.04	0.25	0.07	0.00	0.00	0.36
	30–43	0.00	0.07	0.26	0.04	0.00	0.38
<b>Alternate Form</b>	44–52	0.00	0.00	0.04	0.09	0.02	0.15
	53–65	0.00	0.00	0.00	0.02	0.03	0.05
<b>Estimated Proportion Consistently Classified: Total = 0.66, Proficient &amp; Above = 0.91</b>							

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

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
<b>Decision Accuracy</b>	0–23	0.11	0.03	0.00	0.00	0.00	0.15
	24–36	0.02	0.19	0.03	0.00	0.00	0.24
	37–51	0.00	0.03	0.25	0.03	0.00	0.32
<b>All-forms Average</b>	52–60	0.00	0.00	0.04	0.14	0.03	0.21
	61–75	0.00	0.00	0.00	0.03	0.06	0.08
<b>Estimated Proportion Correctly Classified: Total = 0.75, Proficient &amp; Above = 0.93</b>							
<b>Decision Consistency</b>	0–23	0.10	0.04	0.00	0.00	0.00	0.15
	24–36	0.03	0.16	0.04	0.00	0.00	0.24
	37–51	0.00	0.05	0.23	0.05	0.00	0.32
<b>Alternate Form</b>	52–60	0.00	0.00	0.05	0.12	0.04	0.21
	61–75	0.00	0.00	0.00	0.03	0.05	0.08
<b>Estimated Proportion Consistently Classified: Total = 0.66, Proficient &amp; Above = 0.90</b>							

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

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
<b>Decision Accuracy</b>	0–26	0.22	0.05	0.01	0.00	0.00	0.28
	27–33	0.03	0.11	0.04	0.00	0.00	0.19
	34–46	0.00	0.03	0.22	0.01	0.00	0.25
<b>All-forms Average</b>	47–57	0.00	0.01	0.04	0.13	0.01	0.19
	58–75	0.00	0.00	0.00	0.02	0.06	0.09
<b>Estimated Proportion Correctly Classified: Total = 0.73, Proficient &amp; Above = 0.93</b>							
<b>Decision Consistency</b>	0–26	0.21	0.05	0.01	0.00	0.00	0.28
	27–33	0.05	0.08	0.05	0.00	0.00	0.19
	34–46	0.00	0.04	0.18	0.03	0.00	0.25
<b>Alternate Form</b>	47–57	0.00	0.01	0.05	0.11	0.02	0.19
	58–75	0.00	0.00	0.00	0.02	0.06	0.09
<b>Estimated Proportion Consistently Classified: Total = 0.64, Proficient &amp; Above = 0.90</b>							

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

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
<b>Decision Accuracy</b>	0–23	0.14	0.05	0.01	0.01	0.00	0.20
	24–32	0.03	0.15	0.05	0.00	0.00	0.22
	33–45	0.00	0.02	0.22	0.01	0.00	0.26
<b>All-forms Average</b>	46–56	0.00	0.01	0.05	0.15	0.02	0.22
	57–75	0.00	0.00	0.00	0.03	0.08	0.10
<b>Estimated Proportion Correctly Classified: Total = 0.73, Proficient &amp; Above = 0.92</b>							
<b>Decision Consistency</b>	0–23	0.13	0.05	0.01	0.01	0.00	0.20
	24–32	0.04	0.12	0.06	0.00	0.00	0.22
	33–45	0.00	0.04	0.19	0.03	0.00	0.26
<b>Alternate Form</b>	46–56	0.00	0.01	0.06	0.12	0.03	0.22
	57–75	0.00	0.00	0.00	0.03	0.07	0.10
<b>Estimated Proportion Consistently Classified: Total = 0.63, Proficient &amp; Above = 0.89</b>							

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

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
<b>Decision Accuracy</b>	0–22	0.07	0.04	0.00	0.00	0.00	0.12
	23–33	0.02	0.19	0.04	0.00	0.00	0.25
	34–47	0.00	0.04	0.26	0.03	0.00	0.33
<b>All-forms Average</b>	48–58	0.00	0.00	0.04	0.15	0.02	0.21
	59–75	0.00	0.00	0.00	0.02	0.07	0.09
<b>Estimated Proportion Correctly Classified: Total = 0.74, Proficient &amp; Above = 0.93</b>							
<b>Decision Consistency</b>	0–22	0.06	0.05	0.00	0.00	0.00	0.12
	23–33	0.04	0.16	0.05	0.00	0.00	0.25
	34–47	0.00	0.06	0.22	0.05	0.00	0.33
<b>Alternate Form</b>	48–58	0.00	0.00	0.05	0.13	0.03	0.21
	59–75	0.00	0.00	0.00	0.02	0.07	0.09
<b>Estimated Proportion Consistently Classified: Total = 0.65, Proficient &amp; Above = 0.90</b>							

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

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
<b>Decision Accuracy</b>	0–20	0.05	0.04	0.00	0.00	0.00	0.09
	21–32	0.02	0.21	0.04	0.00	0.00	0.27
	33–46	0.00	0.04	0.26	0.03	0.00	0.33
<b>All-forms Average</b>	47–57	0.00	0.00	0.04	0.15	0.02	0.22
	58–75	0.00	0.00	0.00	0.02	0.07	0.09
<b>Estimated Proportion Correctly Classified: Total = 0.75, Proficient &amp; Above = 0.92</b>							
<b>Decision Consistency</b>	0–20	0.05	0.04	0.00	0.00	0.00	0.09
	21–32	0.04	0.18	0.06	0.00	0.00	0.27
	33–46	0.00	0.06	0.23	0.05	0.00	0.33
<b>Alternate Form</b>	47–57	0.00	0.00	0.05	0.13	0.03	0.22
	58–75	0.00	0.00	0.00	0.02	0.07	0.09
<b>Estimated Proportion Consistently Classified: Total = 0.65, Proficient &amp; Above = 0.90</b>							

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

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
<b>Decision Accuracy</b>	0–22	0.06	0.02	0.00	0.00	0.00	0.09
	23–35	0.02	0.21	0.04	0.00	0.00	0.27
	36–48	0.00	0.04	0.26	0.03	0.00	0.33
<b>All-forms Average</b>	49–58	0.00	0.00	0.05	0.16	0.02	0.22
	59–75	0.00	0.00	0.00	0.03	0.06	0.09
<b>Estimated Proportion Correctly Classified: Total = 0.75, Proficient &amp; Above = 0.92</b>							
<b>Decision Consistency</b>	0–22	0.06	0.03	0.00	0.00	0.00	0.09
	23–35	0.03	0.18	0.06	0.00	0.00	0.27
	36–48	0.00	0.05	0.22	0.05	0.00	0.33
<b>Alternate Form</b>	49–58	0.00	0.00	0.06	0.13	0.03	0.22
	59–75	0.00	0.00	0.00	0.03	0.06	0.09
<b>Estimated Proportion Consistently Classified: Total = 0.65, Proficient &amp; Above = 0.89</b>							

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

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
<b>Decision Accuracy</b>	0–22	0.08	0.03	0.00	0.00	0.00	0.12
	23–36	0.02	0.22	0.03	0.00	0.00	0.27
	37–48	0.00	0.04	0.23	0.04	0.00	0.32
<b>All-forms Average</b>	49–57	0.00	0.00	0.04	0.14	0.02	0.20
	58–75	0.00	0.00	0.00	0.03	0.05	0.09
<b>Estimated Proportion Correctly Classified: Total = 0.73, Proficient &amp; Above = 0.92</b>							
<b>Decision Consistency</b>	0–22	0.08	0.04	0.00	0.00	0.00	0.12
	23–36	0.03	0.19	0.05	0.00	0.00	0.27
	37–48	0.00	0.06	0.19	0.06	0.01	0.32
<b>Alternate Form</b>	49–57	0.00	0.00	0.05	0.12	0.04	0.20
	58–75	0.00	0.00	0.00	0.04	0.05	0.09
<b>Estimated Proportion Consistently Classified: Total = 0.63, Proficient &amp; Above = 0.89</b>							

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

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total
<b>Decision Accuracy</b>	0–22	0.07	0.03	0.01	0.00	0.00	0.11
	23–36	0.00	0.19	0.03	0.00	0.00	0.21
	37–47	0.00	0.04	0.23	0.04	0.00	0.30
<b>All-forms Average</b>	48–56	0.01	0.01	0.06	0.17	0.03	0.28
	57–75	0.00	0.00	0.00	0.03	0.07	0.10
<b>Estimated Proportion Correctly Classified: Total = 0.72, Proficient &amp; Above = 0.88</b>							
<b>Decision Consistency</b>	0–22	0.07	0.03	0.01	0.00	0.00	0.11
	23–36	0.01	0.16	0.04	0.00	0.00	0.21
	37–47	0.00	0.06	0.19	0.06	0.00	0.30
<b>Alternate Form</b>	48–56	0.01	0.01	0.08	0.14	0.05	0.28
	57–75	0.00	0.00	0.00	0.04	0.06	0.10
<b>Estimated Proportion Consistently Classified: Total = 0.62, Proficient &amp; Above = 0.85</b>							

## Appendix 8.C—IRT Analyses

**Table 8.C.1 IRT *b*-values for RLA, Grade Six**

Content Area	No. of items	Mean	Standard Deviation	Minimum	Maximum
Word Analysis and Vocabulary Development	13	0.12	0.59	-0.66	1.27
Reading Comprehension	17	-0.04	0.51	-0.77	0.75
Literary Response and Analysis	12	-0.19	0.44	-0.88	0.49
Written Conventions	16	-0.30	0.58	-1.95	0.44
Writing Strategies	17	0.26	0.42	-0.50	0.94
All operational items	75	-0.02	0.54	-1.95	1.27

**Table 8.C.2 IRT *b*-values for RLA, Grade Eight**

Content Area	No. of items	Mean	Standard Deviation	Minimum	Maximum
Word Analysis and Vocabulary Development	9	-0.39	0.55	-1.15	0.37
Reading Comprehension	18	0.04	0.63	-1.05	1.12
Literary Response and Analysis	15	0.08	0.60	-1.18	0.78
Written Conventions	16	-0.66	0.71	-1.89	0.38
Writing Strategies	17	0.07	0.64	-1.40	1.17
All operational items	75	-0.15	0.69	-1.89	1.17

**Table 8.C.3 Distribution of IRT Difficulty (*b*-values) for RLA (operational items)**

IRT <i>b</i> -value	Grade 6	Grade 8
≥ 3.5	0	0
3.0 < 3.5	0	0
2.5 < 3.0	0	0
2.0 < 2.5	0	0
1.5 < 2.0	0	0
1.0 < 1.5	2	2
0.5 < 1.0	9	10
0.0 < 0.5	25	23
-0.5 < 0.0	23	19
-1.0 < -0.5	15	10
-1.5 < -1.0	0	8
-2.0 < -1.5	1	3
-2.5 < -2.0	0	0
-3.0 < -2.5	0	0
-3.5 < -3.0	0	0
< -3.5	0	0
Total	75	75

## Scaling and Equating Results

**Table 8.C.4 New Conversion for RLA, Grade Two**

Grade 2									
Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score	Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score
0	—	N/A	150.0000	150	33	45	-0.5094	296.8017	297
1	—	-4.9758	150.0000	150	34	48	-0.4412	300.0635	300
2	—	-4.2540	150.0000	150	35	38	-0.3728	303.3319	303
3	—	-3.8201	150.0000	150	36	41	-0.3042	306.6128	307
4	—	-3.5042	153.6182	154	37	47	-0.2352	309.9120	310
5	—	-3.2531	165.6257	166	38	44	-0.1657	313.2360	313
6	—	-3.0429	175.6740	176	39	43	-0.0955	316.5913	317
7	—	-2.8610	184.3700	184	40	43	-0.0245	319.9849	320
8	—	-2.6998	192.0774	192	41	40	0.0474	323.4244	323
9	—	-2.5544	199.0300	199	42	48	0.1205	326.9178	327
10	11	-2.4214	205.3897	205	43	45	0.1949	330.4740	330
11	—	-2.2983	211.2716	211	44	44	0.2708	334.1027	334
12	—	-2.1835	216.7612	217	45	56	0.3484	337.8145	338
13	13	-2.0755	221.9240	222	46	42	0.4281	341.6250	342
14	18	-1.9733	226.8100	227	47	49	0.5101	345.5442	346
15	31	-1.8761	231.4608	231	48	52	0.5947	349.5901	350
16	22	-1.7830	235.9093	236	49	52	0.6824	353.7816	354
17	32	-1.6936	240.1828	240	50	42	0.7736	358.1411	358
18	29	-1.6074	244.3040	244	51	54	0.8688	362.6955	363
19	35	-1.5240	248.2925	248	52	35	0.9689	367.4772	367
20	23	-1.4430	252.1646	252	53	49	1.0745	372.5263	373
21	34	-1.3642	255.9350	256	54	47	1.1867	377.8924	378
22	34	-1.2872	259.6167	260	55	49	1.3070	383.6439	384
23	30	-1.2118	263.2194	263	56	31	1.4370	389.8603	390
24	35	-1.1379	266.7545	267	57	36	1.5792	396.6578	397
25	39	-1.0652	270.2307	270	58	22	1.7369	404.1954	404
26	35	-0.9935	273.6564	274	59	32	1.9148	412.7042	413
27	38	-0.9228	277.0387	277	60	30	2.1209	422.5576	423
28	39	-0.8528	280.3852	280	61	18	2.3675	434.3475	434
29	44	-0.7834	283.7023	284	62	12	2.6784	449.2106	449
30	34	-0.7145	286.9966	287	63	—	3.1069	469.6975	470
31	44	-0.6459	290.2747	290	64	—	3.8221	503.8919	504
32	39	-0.5776	293.5406	294	65	—	N/A	600.0000	600

Note: Performance-level cut scores are highlighted. To protect student privacy, the frequency distribution is not shown if based on 10 or fewer student records.

**Table 8.C.5 New Conversion for RLA, Grade Three**

Grade 3									
Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score	Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score
0	—	N/A	150.0000	150	33	36	-0.2104	311.7400	312
1	—	-4.7085	150.0000	150	34	29	-0.1419	315.3116	315
2	—	-3.9837	150.0000	150	35	43	-0.0733	318.8899	319
3	—	-3.5471	150.0000	150	36	39	-0.0044	322.4812	322
4	—	-3.2288	154.3328	154	37	39	0.0648	326.0924	326
5	—	-2.9755	167.5418	168	38	46	0.1346	329.7302	330
6	—	-2.7634	178.6025	179	39	33	0.2050	333.4019	333
7	—	-2.5798	188.1786	188	40	31	0.2762	337.1151	337
8	—	-2.4170	196.6672	197	41	24	0.3484	340.8780	341
9	—	-2.2701	204.3259	204	42	36	0.4217	344.6993	345
10	—	-2.1358	211.3312	211	43	33	0.4963	348.5905	349
11	—	-2.0116	217.8094	218	44	23	0.5724	352.5600	353
12	—	-1.8957	223.8547	224	45	29	0.6503	356.6207	357
13	—	-1.7867	229.5380	230	46	24	0.7301	360.7862	361
14	11	-1.6836	234.9160	235	47	33	0.8123	365.0718	365
15	16	-1.5854	240.0336	240	48	25	0.8971	369.4954	369
16	23	-1.4916	244.9270	245	49	21	0.9850	374.0776	374
17	28	-1.4015	249.6263	250	50	20	1.0764	378.8428	379
18	30	-1.3146	254.1568	254	51	17	1.1718	383.8199	384
19	43	-1.2305	258.5404	259	52	13	1.2720	389.0458	389
20	47	-1.1490	262.7938	263	53	15	1.3778	394.5626	395
21	38	-1.0696	266.9344	267	54	11	1.4903	400.4254	400
22	45	-0.9921	270.9759	271	55	11	1.6107	406.7046	407
23	40	-0.9162	274.9306	275	56	—	1.7408	413.4919	413
24	43	-0.8418	278.8100	279	57	—	1.8831	420.9110	421
25	34	-0.7687	282.6228	283	58	—	2.0409	429.1401	429
26	47	-0.6967	286.3793	286	59	—	2.2190	438.4288	438
27	45	-0.6256	290.0878	290	60	—	2.4250	449.1730	449
28	45	-0.5552	293.7560	294	61	—	2.6716	462.0312	462
29	39	-0.4855	297.3913	297	62	—	2.9825	478.2442	478
30	38	-0.4163	301.0006	301	63	—	3.4109	500.5818	501
31	50	-0.3475	304.5909	305	64	—	4.1265	537.9014	538
32	44	-0.2789	308.1685	308	65	—	N/A	600.0000	600

Note: Performance-level cut scores are highlighted. To protect student privacy, the frequency distribution is not shown if based on 10 or fewer student records.

**Table 8.C.6 New Conversion for RLA, Grade Four**

Grade 4									
Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score	Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score
0	—	N/A	150.0000	150	38	22	-0.2025	303.5189	304
1	—	-4.7575	150.0000	150	39	14	-0.1434	306.6961	307
2	—	-4.0436	150.0000	150	40	16	-0.0842	309.8804	310
3	—	-3.6173	150.0000	150	41	13	-0.0247	313.0763	313
4	—	-3.3085	150.0000	150	42	23	0.0350	316.2880	316
5	—	-3.0643	150.0000	150	43	12	0.0952	319.5201	320
6	—	-2.8608	160.6341	161	44	17	0.1557	322.7772	323
7	—	-2.6853	170.0653	170	45	24	0.2169	326.0642	326
8	—	-2.5304	178.3941	178	46	15	0.2787	329.3862	329
9	—	-2.3910	185.8830	186	47	13	0.3413	332.7487	333
10	—	-2.2640	192.7117	193	48	22	0.4047	336.1573	336
11	—	-2.1468	199.0084	199	49	15	0.4691	339.6199	340
12	—	-2.0378	204.8679	205	50	20	0.5346	343.1414	343
13	—	-1.9356	210.3631	210	51	18	0.6014	346.7301	347
14	—	-1.8391	215.5488	216	52	18	0.6695	350.3943	350
15	—	-1.7476	220.4711	220	53	23	0.7393	354.1433	354
16	—	-1.6602	225.1660	225	54	21	0.8108	357.9875	358
17	—	-1.5765	229.6634	230	55	19	0.8843	361.9386	362
18	19	-1.4961	233.9880	234	56	17	0.9601	366.0101	366
19	14	-1.4185	238.1609	238	57	14	1.0383	370.2173	370
20	15	-1.3433	242.1999	242	58	16	1.1194	374.5780	375
21	18	-1.2704	246.1203	246	59	22	1.2038	379.1125	379
22	16	-1.1994	249.9361	250	60	16	1.2919	383.8486	384
23	15	-1.1302	253.6578	254	61	—	1.3843	388.8129	389
24	13	-1.0625	257.2968	257	62	16	1.4816	394.0428	394
25	13	-0.9961	260.8621	261	63	—	1.5846	399.5832	400
26	16	-0.9310	264.3628	264	64	—	1.6945	405.4904	405
27	11	-0.8670	267.8044	268	65	—	1.8126	411.8363	412
28	—	-0.8039	271.1953	271	66	—	1.9406	418.7150	419
29	13	-0.7416	274.5418	275	67	—	2.0810	426.2611	426
30	16	-0.6801	277.8496	278	68	—	2.2370	434.6465	435
31	16	-0.6192	281.1242	281	69	—	2.4135	444.1367	444
32	20	-0.5588	284.3705	284	70	—	2.6182	455.1378	455
33	15	-0.4988	287.5937	288	71	—	2.8635	468.3236	468
34	19	-0.4392	290.7982	291	72	—	3.1736	484.9942	485
35	19	-0.3798	293.9887	294	73	—	3.6013	507.9807	508
36	14	-0.3207	297.1695	297	74	—	4.3166	546.4295	546
37	11	-0.2616	300.3448	300	75	—	N/A	600.0000	600

Note: Performance-level cut scores are highlighted. To protect student privacy, the frequency distribution is not shown if based on 10 or fewer student records.

**Table 8.C.7 New Conversion for RLA, Grade Five**

Grade 5									
Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score	Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score
0	—	N/A	150.0000	150	38	13	0.0344	317.1451	317
1	—	-4.4641	150.0000	150	39	13	0.0915	320.9754	321
2	—	-3.7516	150.0000	150	40	13	0.1486	324.8138	325
3	—	-3.3266	150.0000	150	41	12	0.2060	328.6628	329
4	—	-3.0192	150.0000	150	42	11	0.2635	332.5285	333
5	—	-2.7762	150.0000	150	43	11	0.3215	336.4166	336
6	—	-2.5740	150.0000	150	44	12	0.3798	340.3327	340
7	—	-2.3999	153.7267	154	45	12	0.4386	344.2827	344
8	—	-2.2463	164.0404	164	46	12	0.4981	348.2730	348
9	—	-2.1082	173.3052	173	47	13	0.5582	352.3101	352
10	—	-1.9825	181.7451	182	48	14	0.6191	356.4010	356
11	—	-1.8667	189.5201	190	49	11	0.6810	360.5565	361
12	—	-1.7590	196.7490	197	50	11	0.7440	364.7808	365
13	—	-1.6582	203.5210	204	51	12	0.8081	369.0848	369
14	—	-1.5630	209.9071	210	52	16	0.8735	373.4789	373
15	—	-1.4728	215.9630	216	53	17	0.9405	377.9743	378
16	11	-1.3868	221.7340	222	54	—	1.0092	382.5839	383
17	11	-1.3046	227.2572	227	55	—	1.0798	387.3222	387
18	—	-1.2255	232.5635	233	56	—	1.1525	392.2053	392
19	15	-1.1493	237.6794	238	57	—	1.2277	397.2514	397
20	18	-1.0756	242.6261	243	58	—	1.3056	402.4869	402
21	17	-1.0042	247.4237	247	59	—	1.3868	407.9327	408
22	13	-0.9347	252.0886	252	60	—	1.4715	413.6215	414
23	21	-0.8670	256.6353	257	61	—	1.5604	419.5900	420
24	20	-0.8008	261.0767	261	62	—	1.6542	425.8832	426
25	15	-0.7360	265.4244	265	63	—	1.7536	432.5566	433
26	18	-0.6725	269.6886	270	64	—	1.8597	439.6797	440
27	18	-0.6101	273.8788	274	65	—	1.9738	447.3412	447
28	18	-0.5487	278.0034	278	66	—	2.0978	455.6673	456
29	17	-0.4881	282.0709	282	67	—	2.2340	464.8073	465
30	20	-0.4282	286.0870	286	68	—	2.3856	474.9887	475
31	18	-0.3691	290.0596	290	69	—	2.5577	486.5383	487
32	14	-0.3105	293.9949	294	70	—	2.7576	499.9598	500
33	16	-0.2523	297.8990	298	71	—	2.9982	516.1146	516
34	16	-0.1945	301.7776	302	72	—	3.3031	536.5805	537
35	12	-0.1370	305.6361	306	73	—	3.7253	564.9216	565
36	—	-0.0798	309.4800	309	74	—	4.4351	600.0000	600
37	17	-0.0227	313.3148	313	75	—	N/A	600.0000	600

Note: Performance-level cut scores are highlighted. To protect student privacy, the frequency distribution is not shown if based on 10 or fewer student records.

**Table 8.C.8 New Conversion for RLA, Grade Six**

Grade 6									
Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score	Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score
0	—	N/A	150.0000	150	38	11	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	13	0.3492	344.0953	344
7	—	-2.4174	150.0000	150	45	—	0.4080	348.2257	348
8	—	-2.2642	160.3633	160	46	15	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	12	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	13	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	18	-1.0283	247.2503	247	59	—	1.3567	414.9261	415
22	14	-0.9593	252.1043	252	60	—	1.4416	420.8970	421
23	14	-0.8920	256.8354	257	61	—	1.5308	427.1624	427
24	11	-0.8262	261.4572	261	62	—	1.6247	433.7696	434
25	11	-0.7619	265.9819	266	63	—	1.7244	440.7765	441
26	11	-0.6987	270.4201	270	64	—	1.8308	448.2563	448
27	13	-0.6367	274.7818	275	65	—	1.9452	456.3020	456
28	14	-0.5756	279.0771	279	66	—	2.0696	465.0445	465
29	13	-0.5154	283.3112	283	67	—	2.2061	474.6420	475
30	11	-0.4559	287.4939	287	68	—	2.3582	485.3321	485
31	—	-0.3970	291.6321	292	69	—	2.5306	497.4573	497
32	12	-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	12	-0.1661	307.8673	308	73	—	3.7003	579.6889	580
36	14	-0.1091	311.8763	312	74	—	4.4107	600.0000	600
37	11	-0.0522	315.8766	316	75	—	N/A	600.0000	600

Note: Performance-level cut scores are highlighted. To protect student privacy, the frequency distribution is not shown if based on 10 or fewer student records.

**Table 8.C.9 New Conversion for RLA, Grade Seven**

Grade 7									
Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score	Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score
0	—	N/A	150.0000	150	38	—	-0.1211	314.6700	315
1	—	-4.7216	150.0000	150	39	15	-0.0611	318.4072	318
2	—	-4.0063	150.0000	150	40	—	-0.0011	322.1505	322
3	—	-3.5785	150.0000	150	41	12	0.0592	325.9049	326
4	—	-3.2684	150.0000	150	42	11	0.1197	329.6754	330
5	—	-3.0227	150.0000	150	43	12	0.1805	333.4672	333
6	—	-2.8178	150.0000	150	44	12	0.2417	337.2856	337
7	—	-2.6410	157.5875	158	45	—	0.3035	341.1362	341
8	—	-2.4847	167.3306	167	46	—	0.3659	345.0248	345
9	—	-2.3440	176.0998	176	47	12	0.4290	348.9577	349
10	—	-2.2156	184.1031	184	48	—	0.4929	352.9414	353
11	—	-2.0971	191.4891	191	49	—	0.5578	356.9854	357
12	—	-1.9868	198.3679	198	50	—	0.6237	361.0947	361
13	—	-1.8832	204.8238	205	51	13	0.6908	365.2790	365
14	—	-1.7854	210.9209	211	52	—	0.7593	369.5481	370
15	—	-1.6925	216.7121	217	53	—	0.8293	373.9125	374
16	—	-1.6038	222.2392	222	54	—	0.9010	378.3844	378
17	—	-1.5189	227.5368	228	55	—	0.9747	382.9771	383
18	—	-1.4371	232.6336	233	56	—	1.0506	387.7063	388
19	—	-1.3582	237.5538	238	57	—	1.1289	392.5896	393
20	—	-1.2817	242.3181	242	58	11	1.2101	397.6468	398
21	—	-1.2075	246.9448	247	59	—	1.2944	402.9054	403
22	14	-1.1353	251.4478	251	60	—	1.3824	408.3906	408
23	—	-1.0648	255.8419	256	61	—	1.4746	414.1382	414
24	—	-0.9959	260.1390	260	62	—	1.5717	420.1901	420
25	—	-0.9283	264.3495	264	63	—	1.6745	426.5982	427
26	12	-0.8620	268.4830	268	64	—	1.7840	433.4276	433
27	—	-0.7968	272.5482	273	65	—	1.9017	440.7612	441
28	12	-0.7325	276.5536	277	66	—	2.0292	448.7077	449
29	11	-0.6691	280.5049	281	67	—	2.1690	457.4265	457
30	11	-0.6065	284.4100	284	68	—	2.3244	467.1118	467
31	13	-0.5445	288.2750	288	69	—	2.5002	478.0731	478
32	—	-0.4830	292.1058	292	70	—	2.7041	490.7798	491
33	15	-0.4220	295.9080	296	71	—	2.9487	506.0266	506
34	15	-0.3614	299.6868	300	72	—	3.2576	525.2856	525
35	12	-0.3011	303.4475	303	73	—	3.6841	551.8691	552
36	12	-0.2410	307.1950	307	74	—	4.3981	596.3808	596
37	—	-0.1810	310.9343	311	75	—	N/A	600.0000	600

Note: Performance-level cut scores are highlighted. To protect student privacy, the frequency distribution is not shown if based on 10 or fewer student records.

**Table 8.C.10 New Conversion for RLA, Grade Eight**

Grade 8									
Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score	Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score
0	—	N/A	150.0000	150	38	—	−0.1043	318.8006	319
1	—	−4.6982	150.0000	150	39	20	−0.0452	322.5068	323
2	—	−3.9826	150.0000	150	40	11	0.0138	326.2168	326
3	—	−3.5545	150.0000	150	41	11	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	12	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	12	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	15	0.6927	368.8518	369
14	—	−1.7583	214.9152	215	52	12	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	14	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	14	−1.0385	260.1213	260	61	—	1.4606	417.0827	417
24	—	−0.9699	264.4300	264	62	—	1.5558	423.0639	423
25	13	−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	14	−0.5835	288.7007	289	68	—	2.2961	469.5558	470
31	11	−0.5221	292.5549	293	69	—	2.4696	480.4521	480
32	—	−0.4614	296.3721	296	70	—	2.6710	493.1014	493
33	15	−0.4011	300.1579	300	71	—	2.9130	508.3031	508
34	—	−0.3412	303.9178	304	72	—	3.2193	527.5408	528
35	13	−0.2817	307.6572	308	73	—	3.6430	554.1516	554
36	11	−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. To protect student privacy, the frequency distribution is not shown if based on 10 or fewer student records.

**Table 8.C.11 New Conversion for RLA, Grade Nine**

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

Note: Performance-level cut scores are highlighted. To protect student privacy, the frequency distribution is not shown if based on 10 or fewer student records.

**Table 8.C.12 New Conversion for RLA, Grade Ten**

Grade 10									
Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score	Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score
0	—	N/A	150.0000	150	38	11	-0.2020	305.2773	305
1	—	-4.8603	150.0000	150	39	—	-0.1412	309.3009	309
2	—	-4.1423	150.0000	150	40	—	-0.0803	313.3302	313
3	—	-3.7117	150.0000	150	41	—	-0.0193	317.3694	317
4	—	-3.3990	150.0000	150	42	—	0.0420	321.4266	321
5	—	-3.1508	150.0000	150	43	12	0.1036	325.5045	326
6	—	-2.9435	150.0000	150	44	11	0.1656	329.6098	330
7	—	-2.7643	150.0000	150	45	12	0.2281	333.7483	334
8	—	-2.6057	150.0000	150	46	—	0.2912	337.9264	338
9	—	-2.4628	155.6119	156	47	14	0.3550	342.1506	342
10	—	-2.3323	164.2502	164	48	13	0.4196	346.4278	346
11	—	-2.2118	172.2274	172	49	—	0.4852	350.7656	351
12	—	-2.0995	179.6606	180	50	—	0.5518	355.1754	355
13	—	-1.9941	186.6398	187	51	—	0.6196	359.6621	360
14	—	-1.8945	193.2324	193	52	—	0.6887	364.2374	364
15	—	-1.7999	199.4954	199	53	—	0.7593	368.9125	369
16	—	-1.7096	205.4736	205	54	12	0.8316	373.7002	374
17	—	-1.6231	211.2037	211	55	—	0.9059	378.6147	379
18	—	-1.5398	216.7164	217	56	—	0.9823	383.6722	384
19	—	-1.4594	222.0378	222	57	—	1.0611	388.8914	389
20	—	-1.3816	227.1899	227	58	—	1.1427	394.2940	394
21	—	-1.3060	232.1919	232	59	—	1.2275	399.9042	400
22	—	-1.2325	237.0610	237	60	—	1.3159	405.7573	406
23	—	-1.1607	241.8100	242	61	—	1.4084	411.8842	412
24	—	-1.0906	246.4534	246	62	—	1.5058	418.3313	418
25	—	-1.0219	251.0023	251	63	—	1.6089	425.1531	425
26	—	-0.9544	255.4670	255	64	—	1.7186	432.4187	432
27	—	-0.8881	259.8580	260	65	—	1.8364	440.2160	440
28	—	-0.8228	264.1810	264	66	—	1.9640	448.6604	449
29	—	-0.7584	268.4459	268	67	—	2.1038	457.9168	458
30	—	-0.6947	272.6599	273	68	—	2.2591	468.1963	468
31	—	-0.6317	276.8297	277	69	—	2.4347	479.8249	480
32	—	-0.5693	280.9614	281	70	—	2.6383	493.3016	493
33	—	-0.5074	285.0613	285	71	—	2.8825	509.4668	509
34	—	-0.4458	289.1350	289	72	—	3.1910	529.8864	530
35	—	-0.3846	293.1881	293	73	—	3.6168	558.0785	558
36	—	-0.3236	297.2261	297	74	—	4.3301	600.0000	600
37	—	-0.2628	301.2540	301	75	—	N/A	600.0000	600

Note: Performance-level cut scores are highlighted.

**Table 8.C.13 New Conversion for RLA, Grade Eleven**

Grade 11									
Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score	Raw Scr.	Freq. Dist.	Theta	Scale Score	Rprtd Score
0	0	N/A	150.0000	150	38	5	-0.0301	307.6849	308
1	0	-4.6157	150.0000	150	39	8	0.0296	312.1502	312
2	0	-3.8991	150.0000	150	40	4	0.0894	316.6277	317
3	0	-3.4704	150.0000	150	41	7	0.1494	321.1212	321
4	0	-3.1597	150.0000	150	42	6	0.2098	325.6378	326
5	0	-2.9137	150.0000	150	43	4	0.2705	330.1839	330
6	0	-2.7087	150.0000	150	44	7	0.3317	334.7661	335
7	0	-2.5319	150.0000	150	45	3	0.3935	339.3913	339
8	0	-2.3758	150.0000	150	46	1	0.4560	344.0666	344
9	0	-2.2354	150.0000	150	47	5	0.5193	348.7996	349
10	0	-2.1074	152.2257	152	48	11	0.5834	353.6009	354
11	0	-1.9893	161.0584	161	49	3	0.6486	358.4763	358
12	0	-1.8795	169.2780	169	50	10	0.7148	363.4368	363
13	0	-1.7765	176.9837	177	51	3	0.7824	368.4931	368
14	2	-1.6794	184.2565	184	52	2	0.8514	373.6569	374
15	0	-1.5871	191.1591	191	53	8	0.9220	378.9414	379
16	2	-1.4992	197.7425	198	54	3	0.9944	384.3612	384
17	1	-1.4149	204.0487	204	55	10	1.0689	389.9331	390
18	2	-1.3339	210.1126	210	56	1	1.1456	395.6756	396
19	4	-1.2557	215.9637	216	57	2	1.2249	401.6097	402
20	2	-1.1800	221.6277	222	58	4	1.3072	407.7646	408
21	2	-1.1066	227.1248	227	59	1	1.3927	414.1655	414
22	5	-1.0351	232.4751	232	60	5	1.4820	420.8490	421
23	5	-0.9653	237.6949	238	61	3	1.5756	427.8571	428
24	2	-0.8971	242.7987	243	62	1	1.6743	435.2409	435
25	2	-0.8303	247.7996	248	63	0	1.7788	443.0633	443
26	3	-0.7647	252.7093	253	64	2	1.8902	451.4031	451
27	4	-0.7002	257.5384	258	65	0	2.0099	460.3609	460
28	3	-0.6366	262.2973	262	66	0	2.1398	470.0777	470
29	2	-0.5738	266.9930	267	67	0	2.2820	480.7233	481
30	2	-0.5118	271.6352	272	68	0	2.4401	492.5527	493
31	1	-0.4504	276.2314	276	69	0	2.6189	505.9327	506
32	3	-0.3895	280.7889	281	70	0	2.8259	521.4269	521
33	3	-0.3290	285.3145	285	71	0	3.0741	540.0018	540
34	3	-0.2689	289.8146	290	72	0	3.3870	563.4171	563
35	3	-0.2090	294.2958	294	73	0	3.8177	595.6507	596
36	3	-0.1493	298.7640	299	74	0	4.5366	600.0000	600
37	5	-0.0897	303.2254	303	75	0	N/A	600.0000	600

Note: Performance-level cut scores are highlighted. To protect student privacy, the frequency distribution is not shown if based on 10 or fewer student records.

## 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, 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 STS prior to the 2014 administration is described in detail in Chapter 3, starting on page 58. 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 STS administration.

#### Item Specifications

ETS maintained item specifications for each STS 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 CDE. Adherence to the specifications ensured the maintenance of quality and consistency in the item development process.

#### Item Writers

The items for each STS 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 ARPs were committees that advised the CDE and ETS on areas related to item development for the STS. 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 61 in Chapter 3 for additional information on the function of ARPs within the item-review process.

### **Statewide Pupil Assessment Review Panel Review**

The 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 *EC* Section 60602. If the SPAR panel rejected specific items, the items were replaced with other items. See page 64 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 administered in 2014 were statistically reviewed in data review meetings the year they were originally administered. There was no data review of field-test items in 2014. See Table 8.4 on page 145 for the list of the original administrations of each test administered in 2014.

## 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, 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 materials are delivered to the Pearson scanning and scoring facilities in Iowa City, Iowa. The nonscorable materials, including test booklets, are returned to the Security Processing Department in Pearson's Cedar Rapids, Iowa, facility. ETS and Pearson closely monitor the return of materials. The California Technical Assistance Center (CalTAC) at ETS monitors returns and notifies LEAs that do not return their materials in a timely manner. CalTAC contacts the LEA CAASPP Coordinators and works with them to facilitate the return of the test materials.

### **Processing Test Materials**

Upon receipt of the test materials, Pearson 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.

Pearson'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, Pearson conducts a complete check of the scanning system. ETS and Pearson create test decks for every test and form. Each test deck consists of approximately 25 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 each 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, Pearson 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.

Pearson 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:

1. Grade counts are compared to the District Master File Sheets.
2. Document counts are compared to the School Master File Sheets.
3. Document counts are compared to the SGIDs.

Any discrepancies identified in the steps outlined above are followed up by Pearson 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, 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:

- General Reporting Specifications
- Form Planner Specifications
- Aggregation Rules
- “What If” List
- Edit Specifications
- Reporting Cluster Names and Item Numbers

Each of these documents is explained in detail in Chapter 7, starting on page 102. The scoring specifications are reviewed and revised by the CDE, ETS, and Pearson 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 Pearson. 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 and Pearson take 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 104 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 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, 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 before ETS delivers them to Pearson.

During the current administration, the data derived from previous item analyses are used to pre-equate the 2014 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. For tests with item replacements, the new scoring tables are checked against scoring tables in the reuse year conversion for reasonableness. In addition, prior to reporting in 2014, 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**

Pearson utilizes the raw-to-scale scoring tables to assign scale scores for each student. ETS verifies Pearson’s scale scores by independently generating the scale scores for students in a small number of LEAs and comparing these scores with those generated by Pearson. 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 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 forms that are reused in 2014, 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.

## **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, ETS, and Pearson prior to any LEA mailings

All reports are required to include a single, accurate 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. Pearson 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, Pearson 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 multiple-choice tests were made available to LEAs prior to the printing of paper reports. The Quick-turnaround Reporting module of the Test 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 must successfully process the QC file. Once the data were deemed reliable and Pearson has processed a scorable answer document for every student who took a CMA 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 STS 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 seven over a period of the three most recent years of administration—2012, 2013, and 2014—and the base year. For the STS for RLA in grades eight through eleven, these data will be reported for 2013 (the base year) and 2014.

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 SEM, as well as the mean IRT *b*-value for each STS.

The base year of each STS 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 are presented in Table 10.1.

**Table 10.1 Base Years for STS**

Content Area	STS	Base Year
Reading/Language Arts	2	2009
	3	2009
	4	2009
	5	2010
	6	2010
	7	2010
	8	2013
	9	2013
	10	2013
	11	2013

The base years differ over STS 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 SBE 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 197 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 seven for the base year and in 2012, 2013, and 2014 and for grades eight through eleven for base year 2013 and 2014. As noted in previous chapters, the STS reporting scales range from 150 to 600 for all of the tests.

STS 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 on page 197; please note that this information may differ slightly from information found on the CDE's 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.3 through Table 10.A.5 show the distribution of scale scores observed in the base year, which differs according to test, and subsequent administrations in 2012, 2013, and 2014 for the tests in grades two through seven; and in the base year and 2014 for the tests in grades eight 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 tests of the STS, 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 over the comparison years indicate that the STS 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.

Table 10.B.1 in Appendix 10.B, which starts on page 200, presents the average proportion-correct values for the operational items in each STS 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 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. 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 Across Base Year, 2012, 2013, and 2014**

Content Area	STS	Number of Examinees (valid scores)				Scale Score Mean and Standard Deviation							
		Base	2012	2013	2014	Base		2012		2013		2014	
						Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Reading/ Language Arts	2	13,300	10,628	8,854	1,930	333	53	333	57	329	55	319	55
	3	8,992	6,770	6,259	1,388	333	51	335	55	333	53	312	45
	4	4,897	4,431	3,903	803	325	51	327	53	326	54	316	52
	5	3,049	2,875	2,743	645	318	56	323	63	317	59	313	59
	6	1,962	1,768	1,725	482	318	54	325	61	323	62	318	62
	7	1,587	1,456	1,414	460	322	52	330	56	333	60	323	57
	8	1,083	–	1,083	439	330	58	–	–	330	58	323	57
	9	1,980	–	1,980	956	324	54	–	–	324	54	320	54
	10	1,177	–	1,177	388	327	55	–	–	327	55	314	58
	11	669	–	669	183	315	61	–	–	315	61	319	61

**Table 10.A.2 Percentage of Proficient and Above and Percentage of Advanced Across Base Year, 2012, 2013, and 2014**

Content Area	STS	% Proficient and Above				% Advanced			
		Base	2012	2013	2014	Base	2012	2013	2014
Reading/ Language Arts	2	40%	40%	37%	30%	17%	18%	17%	10%
	3	37%	40%	36%	20%	13%	15%	14%	5%
	4	35%	36%	35%	29%	12%	14%	14%	8%
	5	29%	32%	29%	28%	9%	13%	10%	9%
	6	28%	34%	32%	32%	7%	11%	11%	10%
	7	31%	37%	38%	30%	8%	12%	15%	9%
	8	35%	–	35%	31%	11%	–	11%	9%
	9	33%	–	33%	31%	11%	–	11%	9%
	10	37%	–	37%	29%	11%	–	11%	9%
	11	30%	–	30%	38%	9%	–	9%	10%

**Table 10.A.3 Observed Score Distributions of the STS Across Base Year, 2012, 2013, and 2014 (Grades Two through Four)**

Observed Score Distributions	RLA Grade 2				RLA Grade 3				RLA Grade 4			
	Base	2012	2013	2014	Base	2012	2013	2014	Base	2012	2013	2014
570–600	8	15	5	4	2	6	1	0	0	0	0	0
540–569	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1	1	0	0
510–539	N/A	N/A	N/A	N/A	7	4	7	0	1	N/A	N/A	N/A
480–509	18	47	26	4	23	21	14	0	1	4	12	0
450–479	150	92	136	10	100	118	89	2	19	35	22	3
420–449	385	565	274	60	404	280	276	20	105	100	171	12
390–419	1,320	1,152	847	121	646	556	516	45	364	373	328	43
360–389	2,187	1,545	1,470	234	1,382	1,155	1,080	153	818	747	635	113
330–359	3,203	2,082	1,534	382	2,032	1,462	1,120	255	1,007	907	684	150
300–329	2,416	1,987	1,649	392	1,860	1,324	1,276	318	955	871	766	167
270–299	1,941	1,615	1,484	357	1,587	957	1,097	338	846	659	588	139
240–269	1,231	1,164	1,090	252	812	685	658	225	584	537	537	117
210–239	409	340	313	98	126	193	113	32	181	186	148	53
180–209	29	24	26	15	11	9	12	0	15	11	12	4
150–179	3	0	0	1	0	0	0	0	0	0	0	2

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

**Table 10.A.4 Observed Score Distributions of the STS Across Base Year, 2012, 2013, and 2014 (Grades Five through Seven)**

Observed Score Distributions	RLA Grade 5				RLA Grade 6				RLA Grade 7			
	Base	2012	2013	2014	Base	2012	2013	2014	Base	2012	2013	2014
570–600	0	1	0	0	0	1	1	0	0	0	1	0
540–569	0	2	2	0	0	0	1	0	0	1	1	0
510–539	2	9	3	1	1	4	2	1	0	0	4	0
480–509	10	20	9	1	4	4	12	1	3	5	14	1
450–479	27	58	30	6	11	31	42	9	8	23	34	8
420–449	116	123	83	26	50	88	65	20	49	62	73	19
390–419	186	253	214	39	148	129	125	30	122	131	140	31
360–389	359	284	284	80	207	256	218	70	224	211	169	64
330–359	511	456	463	85	369	303	260	71	251	274	268	76
300–329	602	558	493	106	412	306	336	77	327	275	240	91
270–299	583	479	507	139	388	310	294	84	327	259	238	78
240–269	460	431	452	104	271	215	256	77	208	162	182	71
210–239	169	164	175	50	88	95	97	37	66	52	41	19
180–209	23	37	26	6	13	23	16	5	1	1	9	2
150–179	1	0	2	2	0	3	0	0	1	0	0	0

**Table 10.A.5 Observed Score Distributions of the STS Across Base Year and 2014 for RLA (Grades Eight through Eleven)**

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

## Appendix 10.B—Historical Comparisons Tables, Test Characteristics

**Table 10.B.1 Average Proportion Correct for Operational Test Items Across Base Year, 2012, 2013, and 2013**

Content Area	Grade-level STS	Average <i>p</i> -value			
		Base	2012	2013	2014
Reading/Language Arts	2	0.65	0.64	0.62	0.59
	3	0.59	0.61	0.59	0.50
	4	0.57	0.59	0.58	0.55
	5	0.49	0.52	0.51	0.49
	6	0.48	0.52	0.52	0.50
	7	0.53	0.56	0.58	0.53
	8	0.56	–	0.56	0.52
	9	0.55	–	0.55	0.54
	10	0.57	–	0.57	0.53
	11	0.56	–	0.56	0.55

**Table 10.B.2 Overall IRT *b*-values for Operational Test Items Across Base Year, 2012, 2013, and 2014**

Content Area	Grade-level STS	Mean IRT <i>b</i> -value			
		Base	2012	2013	2014
Reading/Language Arts	2	–0.59	–0.55	–0.52	–0.55
	3	–0.31	–0.39	–0.30	–0.26
	4	–0.18	–0.23	–0.18	–0.23
	5	0.09	–0.02	–0.02	0.00
	6	0.05	–0.03	–0.07	–0.02
	7	–0.16	–0.18	–0.23	–0.16
	8	–0.26	–	–0.26	–0.15
	9	–0.18	–	–0.18	–0.19
	10	–0.25	–	–0.25	–0.24
	11	–0.22	–	–0.22	–0.10

**Table 10.B.3 Average Point-Biserial Correlation for Operational Test Items Across Base Year, 2012, 2013, and 2014**

Content Area	Grade-level STS	Average Point-Biserial Correlation			
		Base	2012	2013	2014
Reading/ Language Arts	2	0.43	0.45	0.44	0.44
	3	0.39	0.41	0.40	0.36
	4	0.39	0.40	0.42	0.40
	5	0.36	0.39	0.37	0.37
	6	0.33	0.37	0.37	0.38
	7	0.35	0.37	0.39	0.37
	8	0.38	–	0.38	0.37
	9	0.37	–	0.37	0.37
	10	0.35	–	0.35	0.37
	11	0.35	–	0.35	0.35

**Table 10.B.4 Score Reliabilities (Cronbach's Alpha) and Standard Error of Measurement (SEM) Across Base Year, 2012, 2013, and 2014**

Content Area	Grade-level STS	Reliability				SEM			
		Base	2012	2013	2014	Base	2012	2013	2014
Reading/ Language Arts	2	0.93	0.94	0.94	0.93	3.34	3.38	3.42	3.49
	3	0.91	0.92	0.92	0.89	3.51	3.47	3.51	3.66
	4	0.93	0.93	0.93	0.93	3.79	3.78	3.79	3.86
	5	0.91	0.92	0.91	0.92	3.96	3.91	3.94	3.95
	6	0.89	0.91	0.92	0.92	3.95	3.95	3.96	3.95
	7	0.90	0.92	0.92	0.92	3.89	3.87	3.84	3.88
	8	0.92	–	0.92	0.92	3.87	–	3.87	3.91
	9	0.91	–	0.91	0.91	3.84	–	3.84	3.82
	10	0.90	–	0.90	0.91	3.87	–	3.87	3.87
	11	0.90	–	0.90	0.90	3.92	–	3.92	3.93