

California Department of Education Standards and Assessment Division



Standards-based Tests in Spanish (STS) Technical Report Spring 2007 Administration

**February 2008
Educational Testing Service (ETS)
Contract no. 5417**

Table of Contents

Chapter 1: Introduction	1
Background	1
Education Code 60602 Legislative Intent	1
Standards-based Tests in Spanish (STS)	2
Overview of the Technical Report	3
Appendix 1.A — 2007 STS Item and Time Chart	4
Chapter 2: STS Development Procedures	5
Test Assembly Procedures	5
Test Specifications	5
Statistical Specifications	5
Content Specifications	5
Item Development	6
Item Review Process	7
Internal Reviews	7
Assessment Review Panels	8
SPAR Panel	9
Technical Characteristics of the Assembled Tests	10
Item Development for 2007	10
Reference	12
Appendix 2.A — Reporting Clusters	13
Reading/Language Arts	13
Mathematics.....	14
Appendix 2.B—Technical Characteristics Tables and Plots	15
Appendix 2.C—Cluster Targets	18
Chapter 3: STS Equating and Scaling Procedures	25
Test Construction and Review	25
Post-Administration Operational Calibration	25
References	27
Appendix 3.A—Raw to Theta Conversion Tables	28
Chapter 4: Validity	30
Validity Evidence Based on Test Content	30
STS Assessment Review Panel.....	30
STS Item Writers	32
STS Development Procedures	33
Validity Evidence Based on Relations to Other Variables	36
References	38
Appendix 4.A—Test Correlations Tables	39
Chapter 5: Score Reports	40
Descriptions of Scores	40
Raw Score.....	40
Percent Correct Score	40
Score Reporting	40
Score Report Applications	40
Score Report Contents	40
Score Distributions and Summary Statistics	41
Reference	42
Appendix 5.A—Types of Score Reports Tables	43
Appendix 5.B—Raw Score Distribution Tables	45
Chapter 6: Test Fairness	46
Demographic Distributions	46
Test Variations, Accommodations and Modifications	48
Category 1: Test Variations	48
Category 2: Accommodations.....	48
Category 3: Modifications	48
Accommodation Summaries	48
DIF Analyses	49
Test Security and Confidentiality	50
ETS’s Office of Testing Integrity (OTI)	50
Test Development.....	51
Item Review.....	51

Item Bank	51
Transfer of Forms and Items to CDE	52
Test Administration	52
Test Delivery	52
Processing and Scoring	53
Transfer of Scores via Secure Data Exchange	53
Statistical Analysis	54
Reporting and Posting Results	54
Student Confidentiality	54
Test Results	54
References	56
Appendix 6.A—Demographic Summary Tables	57
Appendix 6.B—Test Variations, Accommodations, and Modifications	69
Appendix 6.C — Accommodation Summary Tables	71
Appendix 6.D — DIF Tables	84
Chapter 7: Item Quality	85
Item Analyses	85
IRT Analyses	86
Summaries of IRT b-values	86
IRT Model-Data Fit Analyses	87
Summary of Item-Level Analyses	88
References	90
Appendix 7.A—Item-by-Item p-value and Point-Biserial Tables	91
Appendix 7.B—IRT Tables	93
Chapter 8: Reliability	96
Reliability	96
Inter-correlations, Reliabilities, and SEMs for Reporting Clusters	97
Subgroup Reliabilities and SEMs	97
References	98
Appendix 8.A—Reliabilities and SEM Tables	99

Tables

Table 2.1 Target Statistical Specifications for the STS	6
Table 2.B.1 Summary of 2007 STS Projected Technical Characteristics	15
Table 2.B.2 Summary of 2007 STS Projected Statistical Attributes	15
Table 2.C.1 Target Cluster IRT b-values for Reading/Language Arts (Grades 2 thru 4)	18
Table 2.C.2 Target Cluster IRT b-values for Mathematics (Grades 2 thru 4)	18
Table 3.A.1 2007 Raw to Theta Conversions for Reading/Language Arts Grades 2, 3, and 4	28
Table 3.A.2 2007 Raw to Theta Conversions for Mathematics Grades 2, 3, and 4	29
Table 4.1 STS ARP Member Qualifications, by subject and total	31
Table 4.A.1 Correlations Between STS Mathematics Tests and CST Mathematics Tests	39
Table 4.A.2 Correlations Between STS RLA Tests and CST ELA Tests	39
Table 5.1 Mean and Standard Deviation of Raw and Percent Correct Scores for STS Overall Population	41
Table 5.2 Mean and Standard Deviation of Raw and Percent Correct Scores for STS Target Population	41
Table 5.A.1 Score Reports Reflecting STS Results	43
Table 5.B.1 Distribution of Raw Scores for STS Overall Population	45
Table 5.B.2 Distribution of Raw Scores for STS Target Population	45
Table 6.1 Subgroup Definitions	46
Table 6.A.1 Demographic Summary for Reading/Language Arts Grade 2 (Overall Population)	57
Table 6.A.2 Demographic Summary for Reading/Language Arts Grade 2 (Target Population)	58
Table 6.A.3 Demographic Summary for Reading/Language Arts Grade 3 (Overall Population)	59
Table 6.A.4 Demographic Summary for Reading/Language Arts Grade 3 (Target Population)	60
Table 6.A.5 Demographic Summary for Reading/Language Arts Grade 4 (Overall Population)	61
Table 6.A.6 Demographic Summary for Reading/Language Arts Grade 4 (Target Population)	62
Table 6.A.7 Demographic Summary for Mathematics Grade 2 (Overall Population)	63
Table 6.A.8 Demographic Summary for Mathematics Grade 2 (Target Population)	64
Table 6.A.9 Demographic Summary for Mathematics Grade 3 (Overall Population)	65
Table 6.A.10 Demographic Summary for Mathematics Grade 3 (Target Population)	66
Table 6.A.11 Demographic Summary for Mathematics Grade 4 (Overall Population)	67
Table 6.A.12 Demographic Summary for Mathematics Grade 4 (Target Population)	68
Table 6.C.1 Accommodation Summary for Reading/Language Arts Grades 2, 3, and 4	71
Table 6.C.2 Accommodation Summary for Mathematics Grades 2, 3, and 4	77
Table 6.D.1 Field-test Items Exhibiting Significant DIF	84
Table 6.D.2 Male-Female DIF Classifications for Operational Items	84

Table 6.D.3 Male-Female DIF Classifications for Field-test Items.....	84
Table 7.1 Summary of Items and Forms Presented in the 2007 STS	85
Table 7.2 Average and Median Proportion Correct and Point-Biserial	86
Table 7.A.1 Item-by-item p-value and Point-Biserial	91
Table 7.B.1 IRT b-values for Reading/Language Arts Grade 2	93
Table 7.B.2 IRT b-values for Reading/Language Arts Grade 3	93
Table 7.B.3 IRT b-values for Reading/Language Arts Grade 4	93
Table 7.B.4 IRT b-values for Mathematics Grade 2	93
Table 7.B.5 IRT b-values for Mathematics Grade 3	94
Table 7.B.6 IRT b-values for Mathematics Grade 4	94
Table 7.B.7 Distribution of IRT b-values for Operational Items.....	94
Table 7.B.8 Distribution of IRT b-values for Field-test Items	95
Table 7.B.9 IRT Model Data Fit Distribution for Operational Items	95
Table 7.B.10 IRT Model Data Fit Distribution for Field-test Items	95
Table 8.1 Reliabilities and Standard Errors of Measurement for the STS.....	97
Table 8.A.1 Subscore Reliabilities and Correlations for Reading/Language Arts.....	99
Table 8.A.2 Subscore Reliabilities and Correlations for Mathematics.....	99
Table 8.A.3 Reliabilities and SEM for the STS by GENDER.....	100
Table 8.A.4 Reliabilities and SEM for the STS by NSLP	100
Table 8.A.5 Reliabilities and SEM for the STS by SPECIAL EDUCATION.....	100
Table 8.A.6 Reliabilities and SEM for the STS by ATTENDANCE IN U.S. SCHOOLS	100
Table 8.A.7 Reliabilities and SEM for the STS by EL PROGRAM PARTICIPATION	101
Table 8.A.8 Subscore Reliabilities and SEM for Reading/Language Arts by GENDER/NSLP	101
Table 8.A.9 Subscore Reliabilities and SEM for Mathematics by GENDER/NSLP.....	102
Table 8.A.10 Subscore Reliabilities and SEM for Reading/Language Arts by SPECIAL EDUCATION/ATTENDANCE IN U.S. SCHOOLS.....	103
Table 8.A.11 Subscore Reliabilities and SEM for Mathematics by SPECIAL EDUCATION/ATTENDANCE IN U.S. SCHOOLS.....	104
Table 8.A.12 Subscore Reliabilities and SEM for Reading/Language Arts by EL PROGRAM PARTICIPATION	105
Table 8.A.13 Subscore Reliabilities and SEM for Mathematics by EL PROGRAM PARTICIPATION	106

STS Technical Report Acronyms and Initialisms

ARP	Assessment Review Panel
BCLAD	Bilingual, Cross-cultural, Language and Academic Development
CAPA	California Alternate Performance Assessment
CAT/6 Survey	California Achievement Tests, Sixth Edition Survey
CDE	California Department of Education
CLAD	Cross-cultural, Language, and Academic Development
CST	California Standards Tests
DIF	Differential Item Functioning
DPLT	designated primary language test
EL	English Learner
ELA	English–language arts
ELD	English Language Development
EM	expectation maximization
ETS	Educational Testing Service
FTP	file transfer protocol
GENASYS	Generalized Analysis System
IEP	individualized education program
IRT	item response theory
MC	multiple choice
NSLP	National School Lunch Program
OTI	Office of Testing Integrity
p-value	item proportion correct
PEM	Pearson Educational Measurement
Pt-Rbis	point-biserial correlations
RLA	reading/language arts
SBE	State Board of Education
SD	standard deviation
SDAIE	specially designed academic instruction in English
SEM	standard error of measurement
SPAR	statewide pupil assessment review
STAR	Standardized Testing and Reporting
STS	Standards-based Tests in Spanish

Chapter 1: Introduction

Background

In 1997 and 1998, the California State Board of Education (SBE) adopted rigorous content standards in four major content areas: English–language arts (ELA), mathematics, history–social science, 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.

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, was authorized in 1997 by state law (Senate Bill 376). Senate Bill 1448, approved by the Legislature and the Governor in August 2004, reauthorized the STAR Program through January 1, 2011, in grades three through eleven. STAR Program testing in grade two has been extended to the 2007–08 school year (spring 2008 administration).

The primary goal of the STAR Program is to help measure how well students are mastering these content standards. The STAR Program has five components:

- California Standards Tests (CSTs) produced for California public schools
- California Achievement Tests, Sixth Edition Survey (CAT/6 Survey) for grades 3 and 7, published by CTB/McGraw-Hill
- California Alternate Performance Assessment (CAPA), an assessment produced for students with significant cognitive disabilities who are not able to take the CSTs and the CAT/6 Survey
- Standards-based Tests in Spanish (STS), an assessment of California content standards for Spanish-speaking English learners in grades 2–4 in spring 2007¹
- Aprenda: La prueba de logros en español, Tercera edición (Aprenda 3), published by Harcourt Assessment Inc.², for students in grades 5–11

Education Code 60602 Legislative Intent

The results for tests within the STAR Program are used for three primary purposes:

Excerpted from California Education Code Section 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.”

¹ Beginning in spring 2007, *Standards-based Tests in Spanish* will gradually replace Aprenda as the designated primary language test (DPLT) for the STAR Program.

² Beginning in spring 2006, the State Board of Education designated the *Aprenda: La prueba de logros en español, Tercera edición (Aprenda 3)*, to replace the SABLE/2 as the DPLT for the STAR 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.”

Standards-based Tests in Spanish (STS)

Excerpted from California Education Code Section 60640. (f) (1) ...pupils with limited English proficiency who are enrolled in any of grades 2 to 11, inclusive, may take a second achievement test in their primary language. Primary language tests administered pursuant to this subdivision and subdivision (g) shall be subject to the requirements of subdivision (a) of Section 60641. These primary language tests shall produce individual pupil scores that are valid and reliable.”

The purpose of the STS program is to permit students to demonstrate achievement of the California content standards in reading/language arts (RLA) and mathematics through a primary language test in Spanish. These content standards, approved by the State Board of Education, describe what students should know and be able to do at each grade level. The STS test results are not part of the accountability system in California.

STS tests are targeted towards Spanish-speaking English learners who have been in U.S. schools less than a year or who receive instruction in Spanish. However, all students who are English learners 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 been in U.S. schools less than 12 months. These are cumulative, not necessarily consecutive, months. The optional population consists of students who receive instruction in English and who have been in U.S. schools longer than 12 cumulative months.

2007 is the first operational year of the STS. The STS in 2007 included RLA and mathematics for eligible students in grades 2 to 4. Between 8,000 and 18,000 students took each STS in 2007; approximately two-thirds of the test-takers belonged to the STS target population. The details on these statistics for each STS are presented later in this report.

All STS tests are comprised of four-option multiple-choice items. The STS is administered at different times depending upon the progression of the school year within each particular district. Specifically, schools must administer the CSTs, the CAT/6 Survey, the CAPA, and the STS within a 21-day window comprised of the 10 days before and 10 days after the day on which 85 percent of the instructional year is completed.

The STS tests are administered in an untimed fashion. California Department of Education (CDE) guidelines for the time within which most students would be expected to finish the STS by test and grade level can be found in Appendix 1.A — 2007 STS Item and Time Chart.

Results of the STS are reported using percent correct scores. The STS score report information includes percent correct scores at both the test level and the reporting cluster level.

All STS tests include a set of six field-test items that are not counted towards students' scores. The grade-level STS consists of several versions, each of which contains a different set of six field-test items.

Overview of the Technical Report

This technical report describes characteristics of the California Standards-based Tests in Spanish administered in Spring 2007. The report is divided into ten chapters as described below:

- Chapter 1 introduces the STAR Program and STS.
- Chapter 2 describes the procedures followed in developing valid STS items and in constructing the 2007 STS. In addition, characteristics of the constructed 2007 test forms and various reporting clusters of each form are presented in this chapter.
- Chapter 3 documents the STS calibration steps and describes the procedures followed to prepare for equating and scaling the STS in future years. The raw to theta conversions are presented for each STS.
- Chapter 4 details the evidence supporting the validity of the STS.
- Chapter 5 describes the kinds of score reports that are produced at the end of each administration of the STS. It also summarizes the test-level analyses performed on scores obtained on the spring 2007 administration of the tests.
- Chapter 6 highlights the importance of maintaining fairness in the STS for various subgroups. This chapter summarizes demographic differences in performance, describes accommodations and modifications, and reports on analyses of differential item functioning (DIF). This chapter also includes a section describing procedures followed by ETS to ensure test security.
- Chapter 7 discusses the various types of evidence collected to ensure the acceptable quality of operational and field-test items. Summaries of classical item analysis statistics, Rasch difficulty estimates, and evaluations of the Rasch model-data fit are included in this chapter.
- Chapter 8 summarizes the reliability analyses, including test reliability, accuracy, and intercorrelations of reporting cluster scores. Some of these analyses are also replicated for various subsections of the test-taking population in order to look at reliabilities at subgroup levels.

Each chapter contains summary tables in the body of the text. However, extended appendices reporting technical data for the different STS tests are listed at the end of the relevant chapters.

Appendix 1.A — 2007 STS Item and Time Chart

Standards-based Test in Spanish	Grade 2		Grade 3		Grade 4	
	Total No. of Items	Time	Total No. of Items	Time	Total No. of Items	Time
Reading/Language Arts	71	150	71	150	81	170
Part 1		50		50		85
Part 2		50		50		85
Part 3		50		50		
Mathematics	71	150	71	150	71	150
Part 1		50		50		75
Part 2		50		50		75
Part 3		50		50		

Chapter 2: STS Development Procedures

The STS was constructed to measure the California content standards as well as to meet psychometric criteria for difficulty and reliability. The psychometric criteria were evaluated using projections based on item statistics from 2006 fall field-test.

Test Assembly Procedures

Test blueprints for the components of the STAR Program were proposed by Educational Testing Service (ETS), reviewed and approved by the respective Assessment Review Panels (ARPs), also reviewed and approved by the CDE, and presented to the SBE for adoption. For each STS test, the California content standards were used as the basis for choosing items. Additional technical targets (e.g., difficulty and discrimination) for test construction were established based on statistical characteristics of the STS item pool.

Test Specifications

Statistical Specifications

The primary statistical targets used for STS test assembly in 2007 were the test information function based on the item response theory (IRT) item parameters and an average point biserial correlation. When using the Rasch model, the target information function makes it possible to choose items to produce a test that has the desired precision of measurement at all ability levels. The target mean and standard deviation of item b-values consistent with the information curves were also provided to test development staff to help with the test construction process. The point biserial correlation is a measure of how well the items discriminate among test-takers and is related to the overall reliability of the test.

These specifications were developed from the analyses of items field-tested in the fall of 2006; the target values and ranges for the specifications are presented in Table 2.1 on page 6. The minimum target value for an item point-biserial was set at 0.14 for each test.

Assembly targets included the total test target and (reporting) cluster score targets. The latter was used to stabilize cluster score performance across years. The target mean and standard deviations of the IRT b-values for the clusters are presented in Tables 2.C.1 through 2.C.2 in Appendix 2.C on page 18. Since the cluster scores include far fewer items than the total test, greater variability between the target and the constructed curves for the cluster scores are expected. Meeting the target for the total test was of primary importance. The graphics for the information curve of the total test are presented in Figure 2.B.1, starting on page 13, for the RLA and mathematics tests, respectively. These curves present the target test information curves in comparison with the projected test features for the total for each test. Figures 2.C.1 through Figures 2.C.6 present similar information for the cluster scores for the RLA and mathematics tests, respectively.

Content Specifications

ETS developed all STS test items to conform to the SBE-approved content standards and test blueprints. The content blueprints for the STS can be found on the CDE Web site, at <http://www.cde.ca.gov/ta/tg/sr/stsblueprints.asp>. Although the test blueprints called for distributions of items at the individual standard level, for reporting purposes, the content for each STS was aggregated across standards into subcontent areas, referred to as “reporting clusters.” For each STS reporting cluster, the percentage of questions correctly answered by students was reported. A description of the STS reporting clusters and the standards that comprise the reporting clusters is provided in Appendix 2.A — Reporting Clusters starting on page 13.

Table 2.1 Target Statistical Specifications for the STS

Subject	STS	Target Mean b	Target SD b	Min p-value	Max p-value	Mean Point Biserial	Min Point Biserial
<i>Reading/Language Arts</i>	2	-0.45	0.75	0.20	0.95	> 0.37	0.14
	3	-0.48	0.75	0.20	0.95	> 0.37	0.14
	4	-0.44	0.70	0.20	0.95	> 0.37	0.14
<i>Mathematics</i>	2	-0.67	0.83	0.20	0.95	0.39 – 0.45	0.14
	3	-0.61	0.75	0.20	0.95	0.39 – 0.45	0.14
	4	-0.50	0.75	0.20	0.95	0.39 – 0.45	0.14

An important part of the STS specifications for STS is that all items are written in Spanish, rather than translated or transadapted from English. In addition, all commissioned reading passages and all previously published reading passages were originally written in Spanish.

Item Development

ETS has maintained item specifications for each grade-level STS. ETS followed the SBE-approved Item Utilization Plan to guide the development of the quantity of items for each subject area. Item specification documents included the constructs to be measured and the California content standards included in the test blueprints. The item specifications help ensure that the STS tests consistently match the content standards from year to year. Item writing emphasis is determined in consultation with the CDE. The item specifications also provide specific and important guidance to item writers, and ensure that items are consistent in approach and written to measure the standards. The item specifications describe the general characteristics of the items for each content standard, indicate item types, or content to be avoided, and define the content limits for the items. In summary, the specifications include the following:

- A statement of the strand or topic for the standard
- A full statement of the academic content standard, as found in each STS blueprint
- The expected cognitive level(s) of items written for the standard (Acquire, Integrate, Extend), as defined by ETS and approved by CDE
- The construct(s) appropriately measured by the standard
- A description of the kinds of stems appropriate for multiple-choice items for the standard
- A description of the kinds of distracters appropriate for multiple-choice items for the standard
- A description of specific kinds of items to be avoided, if any (e.g., no RLA items about insignificant details)
- A description of appropriate stimuli (e.g., charts, tables, graphs, or other artwork) for mathematics items
- The content limits for the standard (e.g., one or two variables, maximum place values of numbers) for mathematics items
- A description of appropriate reading passages (if applicable) for RLA items

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

- A list of topics to be avoided
- The acceptable ranges for passage length
- The expected distribution of passages by genre
- Guidelines for readability and concept load, using CDE/ETS–agreed-to standards.

- Expected use of artwork
- The target number of items attached to each reading passage and each writing passage

Item Review Process

The items selected for each STS undergo an extensive item review process that is designed to provide all California students with the best standards-based tests possible. This section summarizes the various reviews that contributed to the validity of 2007 STS items and test forms.

Internal Reviews

After the items have been written, ETS employs a series of internal reviews. The reviews establish the criteria used to judge the content validity of the item, making sure that each item is measuring what it is intended to measure. The internal reviews also examine the overall quality of the test items before they are prepared for presentation to CDE and the ARPs. Because of the complexities involved in producing defensible items for high-stakes programs such as the STAR Program, it is essential that many experienced content area assessment specialists review each item before it is brought to CDE and the ARP and Statewide Pupil Assessment Review (SPAR) panels.

The ETS review process for the STS included the following:

- Internal content review
- Internal sensitivity review
- Internal editorial review

Throughout this multi-step item review process, the lead content area assessment specialists and development team members continually evaluate the relevance of the information being assessed, its relevance to the California content standards, its match to the test and item specifications, and its appropriateness to the population being assessed. Items that are only peripherally related to the test and item specifications, that do not measure core outcomes reflected in the California content standards, or that are not developmentally appropriate are eliminated early in this rigorous review process.

Internal Content Review

Test items and materials receive three reviews from the content area assessment specialists which includes a senior content review. These assessment specialists make sure that the test items and related materials are in compliance with ETS's written guidelines for clarity, style, accuracy, and appropriateness for California students, and in compliance with the approved item specifications. Assessment specialists review each item following the criteria below:

- Relevance of each item as the item relates to the purpose of the test
- Match of each item to the item specifications, including cognitive level
- Match of each item to the principles of quality item development
- Match of each item to the identified standard
- 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 artwork, graphs, figures, etc.

The assessment specialists also check all items against their classification codes, both to evaluate the correctness of the classification and to ensure that a given task is of a type appropriate to the outcome it was intended to measure. The reviewers accept the item and classification as written, suggest revisions, or recommend that the item be discarded. These steps occurred prior to CDE review.

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, conduct the next level of review. These trained staff members review every item before it is prepared for CDE and ARP review. In addition, the review process promotes 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

Internal Editorial Review

After the content area assessment specialists review each item, a group of specially trained editors review each item in preparation for review by CDE and the ARPs. The editors check questions 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.

Assessment Review Panels

ETS is responsible for working with ARPs as items are developed for the STS. The ARPs are advisory panels to CDE and ETS on areas related to the item development for the STS. The ARPs are responsible for reviewing all newly developed items for alignment to the California content standards. The ARPs also review the items for accuracy of item content, clarity of phrasing, and item quality. ETS provides the ARPs with the opportunity to review the items with the applicable field-test statistics and to make recommendations for the use of items in subsequent test forms. The ARPs may raise, in their examination of test items, concerns related to age/grade appropriateness, gender, racial/ethnic, and socioeconomic bias.

ARP Meetings for Review of STS Items

The ETS content area assessment specialists facilitate the STS ARP meetings. Each meeting begins with a brief training session on how to review items. ETS provides this training, which consists of the following steps:

- An overview of the purpose and scope of the STS
- An overview of the STS test design specifications and blueprints
- An analysis of the STS item specifications
- An overview of criteria for evaluating multiple-choice test items and for reviewing constructed-response writing tasks
- Review and evaluation of items for bias and sensitivity issues

The criteria for evaluating multiple-choice items and constructed-response writing tasks included:

- 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 distracters
- bias and sensitivity factors

Criteria also include more global issues, including—for RLA—the appropriateness, difficulty, and readability of reading passages. The committee is also trained on how to make recommendations for revising items. Guidelines for reviewing items are provided by ETS and approved by CDE. The set of guidelines for reviewing items is summarized below:

Item Guidelines: 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 from bias against any person or group?

Stimulus Guidelines: 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, panel members review a set of items independently and record their individual comments. The next step in the review process is for the group to discuss each item. The content area assessment specialists facilitate the discussion and record all recommendations. These recommendations are recorded in a master item review booklet. Item review binders and other item evaluation materials also identify potential bias and sensitivity factors the ARP will consider as part of its item reviews. Depending upon CDE approval and the numbers of items still to be reviewed, some ARPs are further divided into smaller groups. These smaller groups are facilitated by the content area assessment specialists, as well.

ETS staff maintain 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.

SPAR Panel

The Statewide Pupil Assessment Review panel is responsible for reviewing and approving achievement tests to be used statewide for the testing of students in California public schools, grades 2–11. At the SPAR panel meetings, all new items are presented in binders for review. The SPAR panel representatives ensure that the test items conformed to the requirements of *Education Code* Section 60614. If the SPAR panel rejects specific items and/or constructed-response writing tasks, the items and/or tasks are replaced. For the SPAR panel meeting, the item development coordinator or the ETS lead content specialist who had been requested in advance by CDE attends the opening

session and remains in a nearby location or near a telephone to be available to respond to any questions during the course of the meeting.

Technical Characteristics of the Assembled Tests

Technical characteristics of the assembled tests prior to administration were estimated using Gulliksen's (1987) formula for estimating test reliability from item p-values and item point-biserial correlations:

$$r_{xx} = \left(\frac{K}{K-1} \right) \left(1 - \frac{\sum_{g=1}^K s_g^2}{\left(\sum_{g=1}^K r_{xg} s_g \right)^2} \right), \quad (2.1)$$

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 means are calculated by summing the item p-values and estimated standard deviations are calculated by summing the item reliability indices. Table 2.B.1 on page 15 presents these summary values by subject area and grade.

In general, the projected reliabilities of the STS were at or above the levels considered acceptable for such tests, ranging from 0.90 to 0.93. It should be noted that the projected reliabilities in Table 2.B.1 were based on item p-values and point-biserial correlations that were based on external field-testing using samples of students that were not fully representative of the state. Chapters 6 and 7 present item p-values, point-biserial correlations, and test reliability estimates based on the data from the 2007 STS administration.

Table 2.B.2 on page 15 shows the mean observed statistics based on field-test statistics for the STS tests. These values can be compared to the target values in Table 2.1. This suggests that the test forms of RLA for grades 3 and 4 and for mathematics for grade 4, based on field-test statistics, were slightly more difficult than the target specifications, which is due to the limitation of the item pool for these tests. The difficulty of the other test forms was in line with the target specifications. The spread of item difficulties was mostly in line with specifications. However, in some cases the observed spread was less than the target spread. The field-test-based projected point-biserial correlation exceeded the target in all cases.

The graphics in Figure 2.B.1 and Figure 2.B.2, for the total test, and Figure 2.C.1 through Figure 2.C.6, for the cluster scores, show the comparisons of the target test information function and the projected test information function based on field test item parameter estimates for each of the STS tests. The projected test information curves were very close to the target curves for most tests, except for RLA grade 3 and 4, for which constructing a test of desired difficulty level from the current pool of items proved to be difficult. It is anticipated that, over the next few administrations, field-testing easier items will continue to adjust the overall difficulty of this test.

Item Development for 2007

As per the 2006 Scope of Work, ETS has developed an Item Utilization Plan to continue the development of items for CST, CMA, CAPA, and STS, over the next five years. This plan includes strategies for continued coverage of all appropriate standards for all tests in each content area and at each grade level.

Item writer training for STS was conducted over two days in Burbank in April 2006. An effort was made to recruit participants who are bilingual/biliterate to participate in STS training. At this session, ETS test development specialists trained attendees in the basics of item writing. They also reviewed items that participants created during the training, offering feedback in both group and individual settings.

The development of new items during this cycle was limited to a level that would allow for replacement of items no longer available for use on operational forms. All item writers met the following minimum qualifications:

- The minimum 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.
- Writers have at least three years of classroom teaching experience at the appropriate grade level.
- 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.
- All STS writers are bilingual and biliterate in Spanish and English.

Reference

Gulliksen, H. *Theory of Mental Tests*. Hillsdale, NJ: Erlbaum, 1987.

Appendix 2.A — Reporting Clusters

Reading/Language Arts

Grade 2 Reading/Language Arts Standards Test

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

Grade 3 Reading/Language Arts Standards Test

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

Grade 4 Reading/Language Arts Standards Test

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

Mathematics**Grade 2 Mathematics Standards Test**

<i>Number Sense</i>	
Place value, addition, and subtraction	15 items
Multiplication, division, and fractions	23 items
<i>Algebra and Functions</i>	
6 items	
<i>Measurement and Geometry</i>	
14 items	
<i>Statistics, Data Analysis, and Probability</i>	
7 items	

Grade 3 Mathematics Standards Test

<i>Number Sense</i>	
Place value, fractions, and decimals	16 items
Addition, subtraction, multiplication, and division	16 items
<i>Algebra and Functions</i>	
12 items	
<i>Measurement and Geometry</i>	
16 items	
<i>Statistics, Data Analysis, and Probability</i>	
5 items	

Grade 4 Mathematics Standards Test

<i>Number Sense</i>	
Decimals, fractions, and negative numbers	17 items
Operations and factoring	14 items
<i>Algebra and Functions</i>	
18 items	
<i>Measurement and Geometry</i>	
12 items	
<i>Statistics, Data Analysis, and Probability</i>	
4 items	

Appendix 2.B—Technical Characteristics Tables and Plots**Table 2.B.1 Summary of 2007 STS Projected Technical Characteristics**

Subject	STS	Number of Items	Mean Raw Score	Standard	Reliability
				Deviation of Raw Scores	
<i>Reading/Language Arts</i>	2	65	38.09	12.17	0.92
	3	65	35.04	11.45	0.90
	4	75	40.08	13.54	0.92
<i>Mathematics</i>	2	65	41.04	11.70	0.92
	3	65	39.95	12.73	0.93
	4	65	37.45	12.12	0.91

Table 2.B.2 Summary of 2007 STS Projected Statistical Attributes

Subject	STS	Mean b	SD b	Mean p-value	Min p-value	Max p-value	Mean Point	Min Point
							Biserial	Biserial
<i>Reading/Language Arts</i>	2	-0.44	0.74	0.59	0.30	0.85	0.40	0.24
	3	-0.22	0.79	0.54	0.28	0.87	0.37	0.20
	4	-0.17	0.66	0.53	0.16	0.80	0.38	0.12
<i>Mathematics</i>	2	-0.65	0.91	0.63	0.28	0.95	0.40	0.20
	3	-0.58	0.71	0.61	0.30	0.89	0.42	0.24
	4	-0.40	0.54	0.58	0.31	0.81	0.39	0.24

Figure 2.B.1 Comparison Plots for Target Information Function and Projected Test Information for Reading/Language Arts

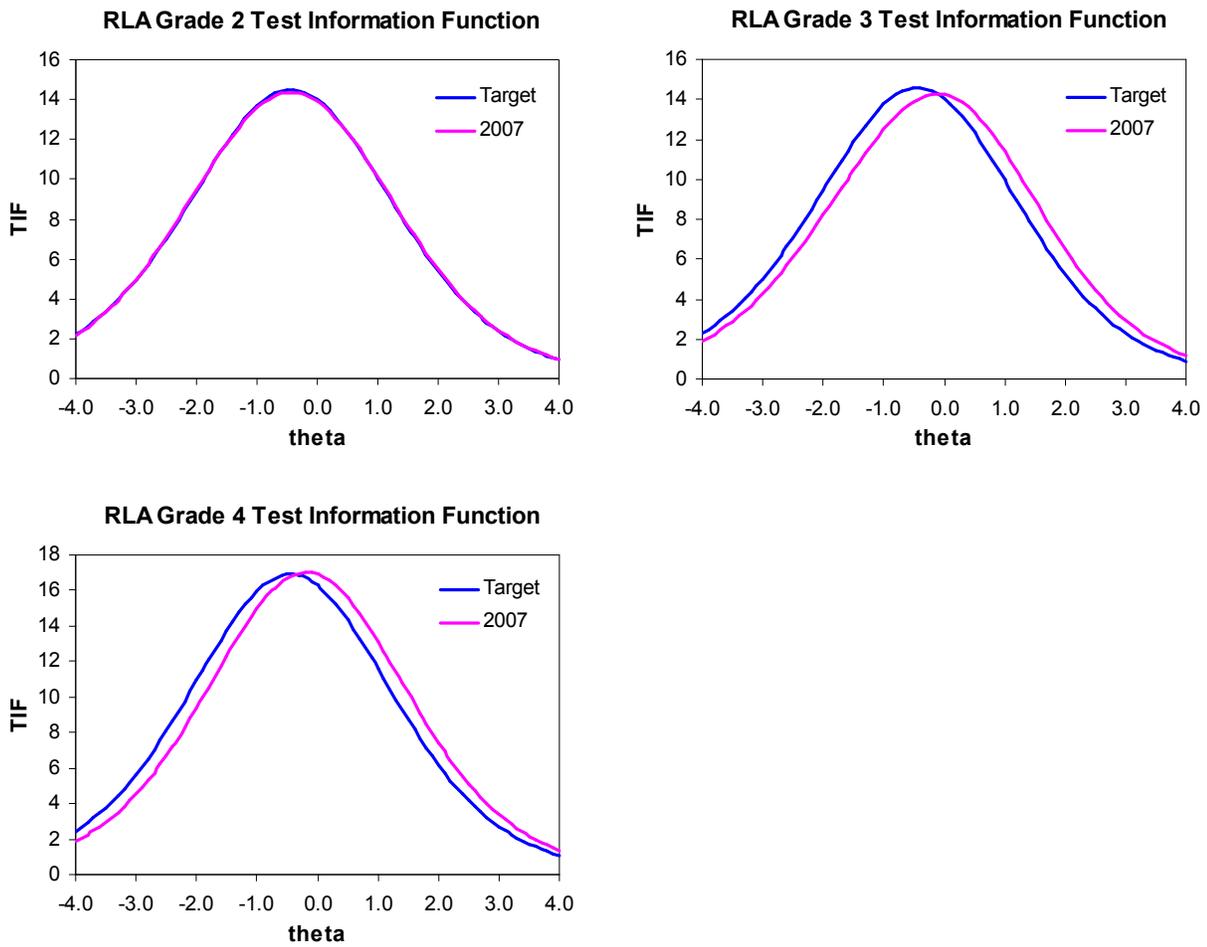
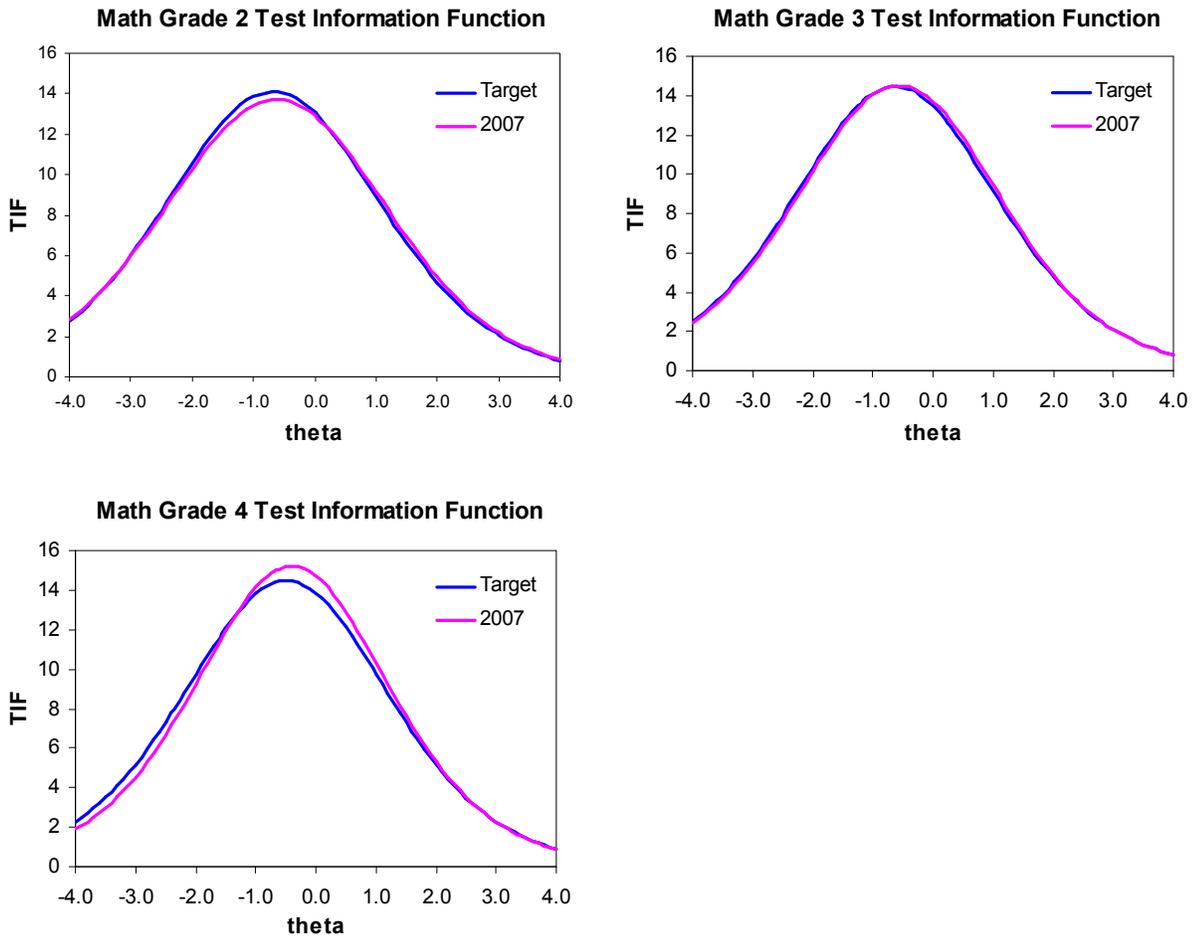


Figure 2.B.2 Comparison Plots for Target Information Function and Projected Test Information for Mathematics



Appendix 2.C—Cluster Targets

Table 2.C.1 Target Cluster IRT b-values for Reading/Language Arts (Grades 2 thru 4)

Clusters	Grade 2			Grade 3			Grade 4		
	N	Mean	Stdev	N	Mean	Stdev	N	Mean	Stdev
1	22	-0.81	0.76	20	-1.06	0.81	18	-0.64	0.78
2	15	-0.38	0.67	15	-0.29	0.60	15	-0.45	0.65
3	6	-0.85	0.87	8	-0.37	0.65	9	-0.24	0.77
4	14	-0.16	0.58	13	-0.26	0.46	18	-0.48	0.66
5	8	0.21	0.37	9	0.09	0.48	15	-0.26	0.67

Table 2.C.2 Target Cluster IRT b-values for Mathematics (Grades 2 thru 4)

Clusters	Grade 2			Grade 3			Grade 4		
	N	Mean	Stdev	N	Mean	Stdev	N	Mean	Stdev
1	15	-0.66	0.79	16	-0.58	0.81	17	-0.55	0.94
2	23	-0.63	0.79	16	-0.64	0.69	14	-0.69	0.73
3	6	-0.26	0.88	12	-0.42	0.77	18	-0.55	0.55
4	14	-0.85	0.98	16	-0.72	0.81	12	-0.21	0.79
5	7	-0.82	0.80	5	-0.72	0.69	4	-0.31	0.61

Figure 2.C.1 Comparison Plots for Target Cluster Information Function and Projected Cluster Information for RLA Grade 2

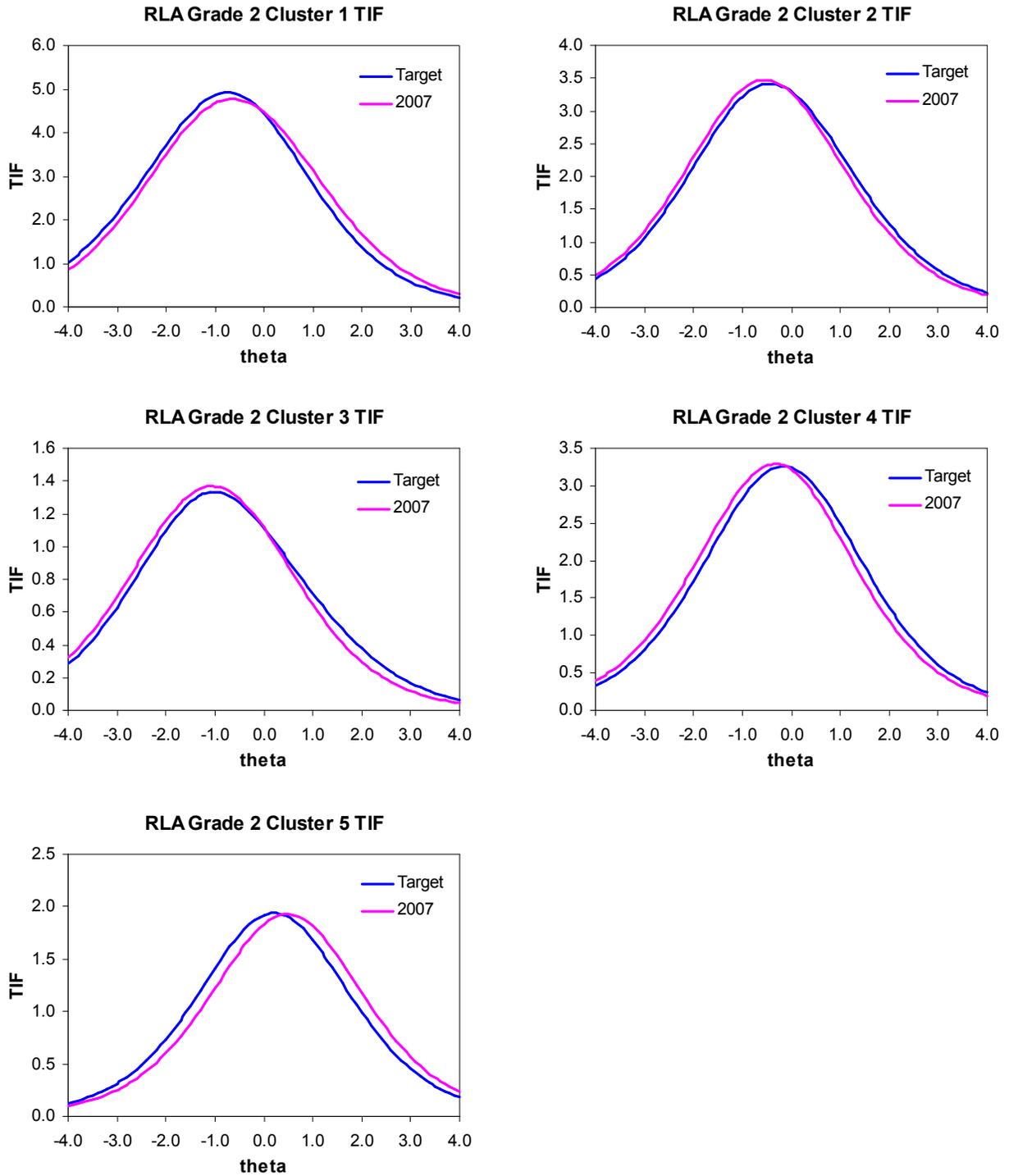


Figure 2.C.2 Comparison Plots for Target Cluster Information Function and Projected Cluster Information for RLA Grade 3

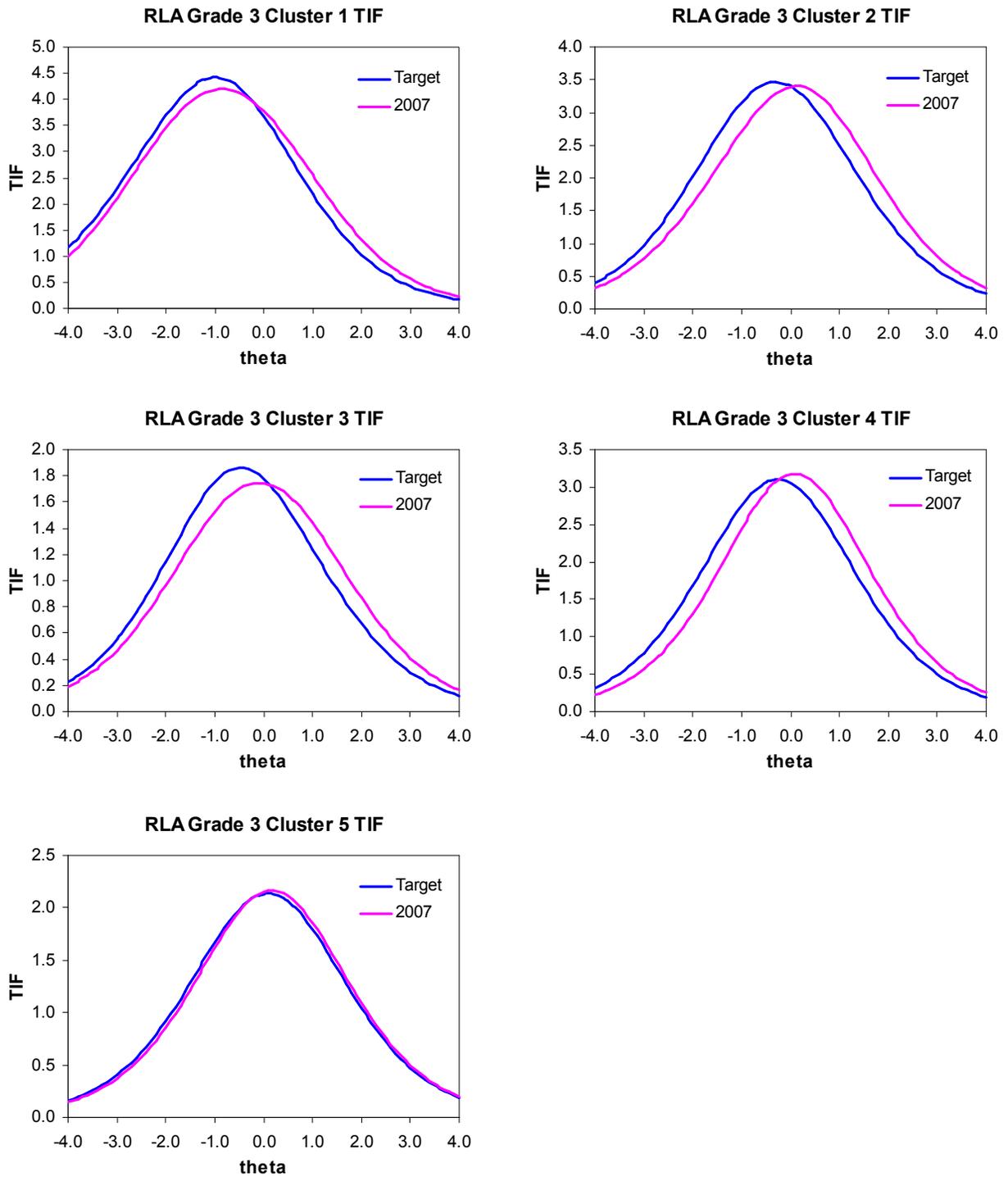


Figure 2.C.3 Comparison Plots for Target Cluster Information Function and Projected Cluster Information for RLA Grade 4

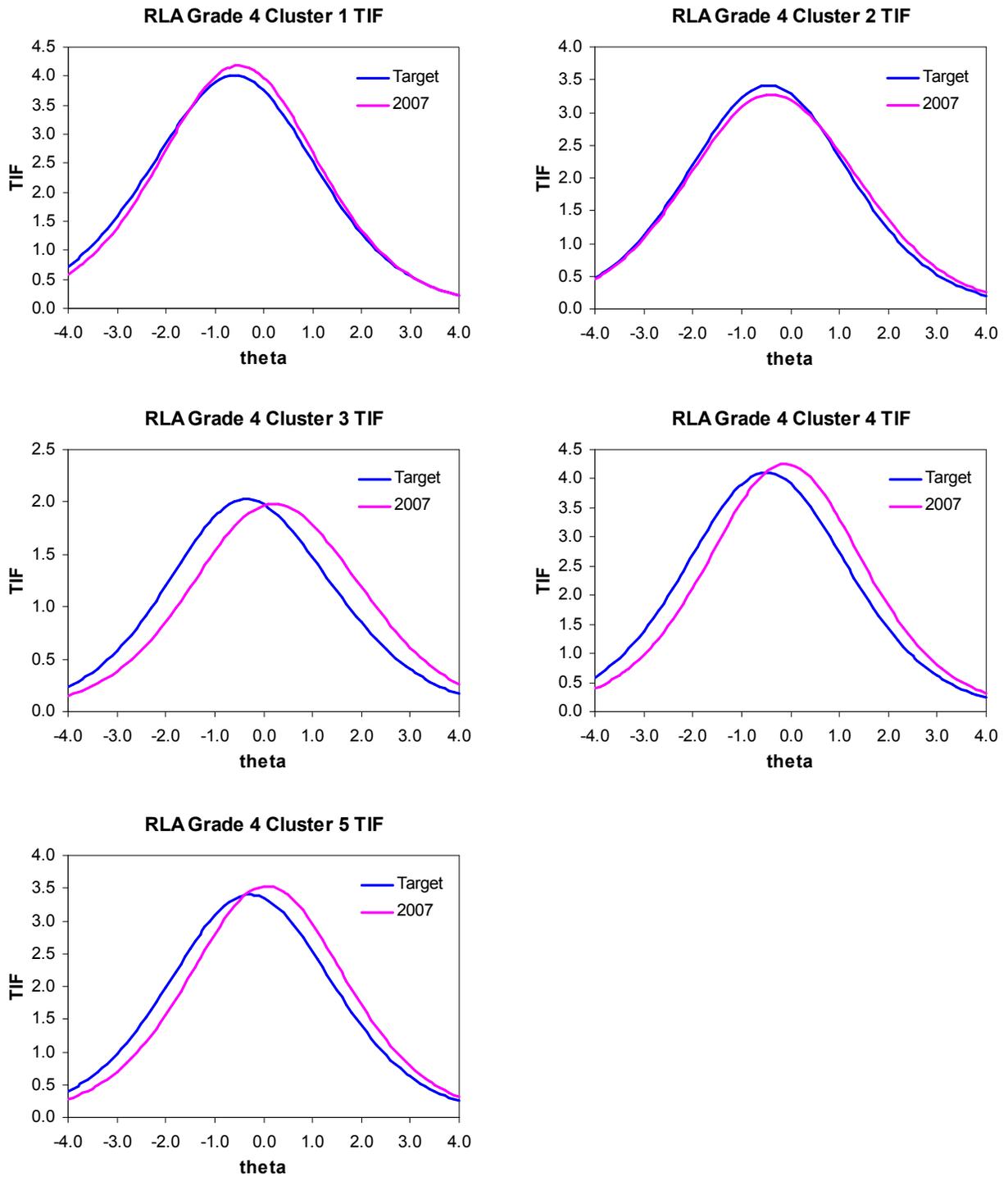


Figure 2.C.4 Comparison Plots for Target Cluster Information Function and Projected Cluster Information for Math Grade 2

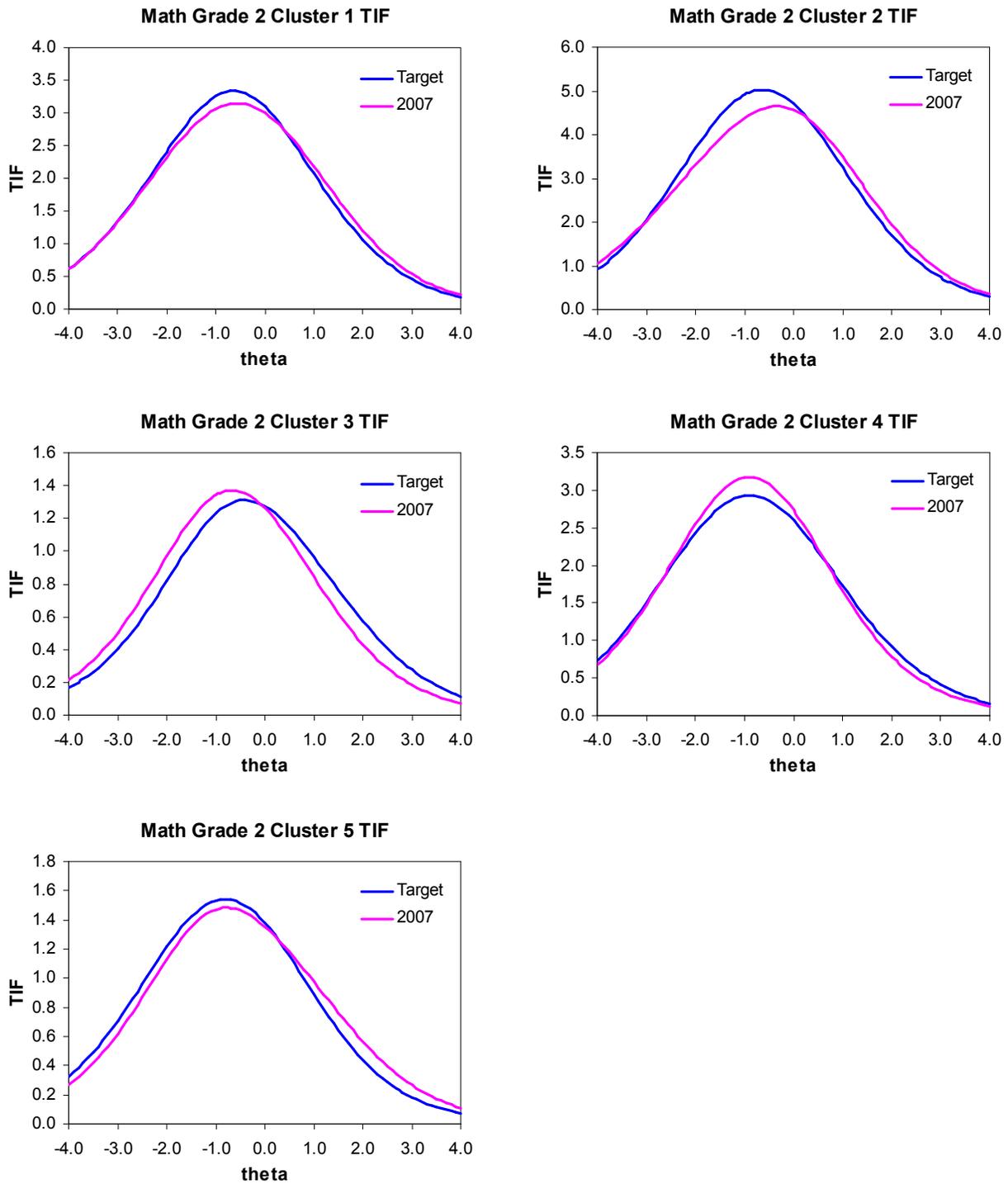


Figure 2.C.5 Comparison Plots for Target Cluster Information Function and Projected Cluster Information for Math Grade 3

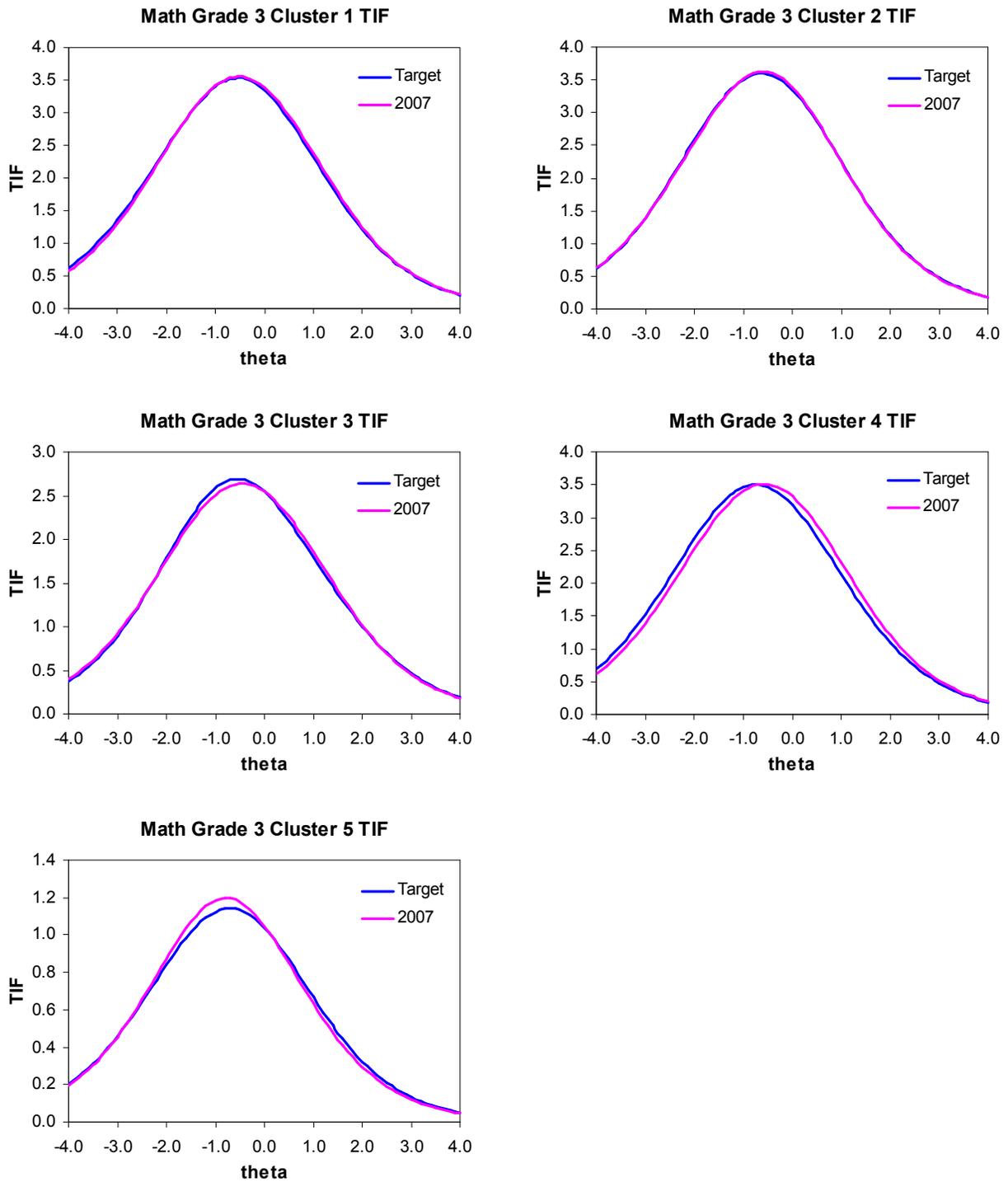
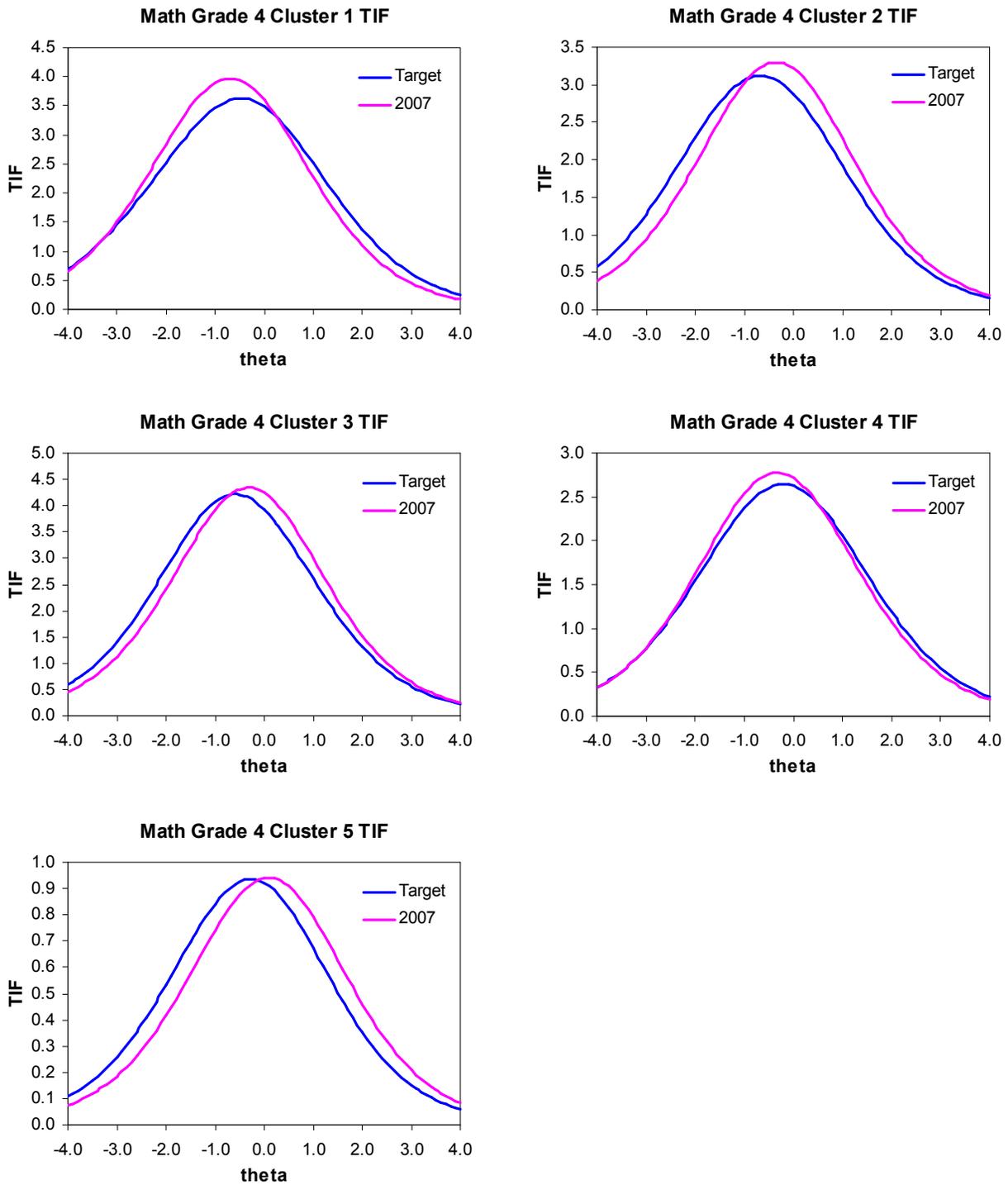


Figure 2.C.6 Comparison Plots for Target Cluster Information Function and Projected Cluster Information for Math Grade 4



Chapter 3: STS Equating and Scaling Procedures

When test forms are created, two primary criteria must be simultaneously satisfied. The first is content-based: Test items must be distributed according to content specifications. The second is statistical: The items must have a specified distribution of difficulty or specified average difficulty and a specified average discrimination (correlation between the item score and the test score). These criteria help assure that all forms of a test are parallel (equally reliable and measure the same construct). However, despite the efforts taken when a test is constructed, forms of a test will still differ in difficulty to a small degree. The equating process is used to adjust for these small differences in difficulty so that test-takers can be fairly compared, regardless of the test form they take. However, because 2007 is the first operational year of STS test, there is no need for the equating procedure. Only item calibration steps were conducted and are described in this chapter.

Test Construction and Review

STS tests are assembled to content and statistical specifications. For the 2007 tests, target test information curves were specified to which the test developers assembled forms. See Figure 2.B.1 and Figure 2.B.2 in Chapter 2 for the relationship between the target and assembled forms.

Post-Administration Operational Calibration

2007 is the first operational year of STS tests, so no equating was conducted. Post-administration operational equating is planned for future test forms using a common-item nonequivalent groups design and methods based on item response theory.

The 2007 form will serve as the reference form. The procedures that will be used for equating the future STS test forms to the 2007 reference form will involve three steps: item calibration, item parameter scaling, and true score equating. The “base” or “reference” calibrations for the STS were established by calibrating all available data of the STS target population from the 2007 administration. This established a scale to which subsequent item calibrations could be linked.

There are two distinct populations that are funded by CDE to take the STS. The first is a set of students who are required to take the test, referred to as the “target population.” The second is a set of student who may optionally take the test and are funded by CDE to do so. The “target population” consists of Spanish-speaking English learners receiving instruction in Spanish or Spanish-speaking English learners who have been in U.S. schools less than 12 cumulative (not consecutive) months. The “optional population” consists of the Spanish-speaking English learners receiving instruction in English who have been in U.S. schools greater than 12 cumulative (not consecutive) months. All item analyses, scaling, and form equating will make use of the target population—the population for whom the STS tests are intended—in the 2007 reference form calibration.

ETS utilizes a computer system called the Generalized Analysis System (GENASYS) for the IRT item calibration and equating work. As part of this system, a proprietary version of the PARSCALE computer program (Muraki and 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, and Julian, 2002).

For the item calibrations, the PARSCALE program is 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 is equivalent to the Rasch model for

multiple-choice items and the Rasch partial credit model for polytomously scored items. For the purposes of score equating, only the operational items will be calibrated for each test.

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

Complete raw-to-theta score conversion tables for the 2007 STS are presented in Tables 3.A.1 and 3.A.2 starting on page 28. These conversion tables, together with the standard setting results, will be used for equating purposes in future operational years.

References

Muraki, E. and Bock, R. D. *PARSCALE: Parameter scaling of rating data* (Version 2.2). Chicago, IL: Scientific Software, Inc., 1995.

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

Appendix 3.A—Raw to Theta Conversion Tables**Table 3.A.1 2007 Raw to Theta Conversions for Reading/Language Arts Grades 2, 3, and 4**

Grade 2				Grade 3				Grade 4			
Raw Score	Theta	Raw Score	Theta	Raw Score	Theta	Raw Score	Theta	Raw Score	Theta	Raw Score	Theta
0	N/A	41	0.0097	0	N/A	41	0.2870	0	N/A	41	0.0299
1	-5.1384	42	0.0882	1	-4.9422	42	0.3625	1	-4.7156	42	0.0898
2	-4.4177	43	0.1682	2	-4.2119	43	0.4392	2	-4.0014	43	0.1501
3	-3.9842	44	0.2498	3	-3.7695	44	0.5174	3	-3.5747	44	0.2109
4	-3.6683	45	0.3332	4	-3.4452	45	0.5973	4	-3.2656	45	0.2722
5	-3.4167	46	0.4187	5	-3.1858	46	0.6791	5	-3.0209	46	0.3341
6	-3.2058	47	0.5067	6	-2.9675	47	0.7631	6	-2.8170	47	0.3968
7	-3.0228	48	0.5974	7	-2.7777	48	0.8497	7	-2.6412	48	0.4603
8	-2.8602	49	0.6912	8	-2.6088	49	0.9394	8	-2.4858	49	0.5249
9	-2.7131	50	0.7886	9	-2.4559	50	1.0324	9	-2.3461	50	0.5905
10	-2.5782	51	0.8902	10	-2.3156	51	1.1295	10	-2.2187	51	0.6574
11	-2.4531	52	0.9967	11	-2.1855	52	1.2313	11	-2.1011	52	0.7257
12	-2.3360	53	1.1088	12	-2.0639	53	1.3387	12	-1.9917	53	0.7956
13	-2.2256	54	1.2275	13	-1.9493	54	1.4526	13	-1.8891	54	0.8672
14	-2.1208	55	1.3544	14	-1.8407	55	1.5744	14	-1.7923	55	0.9408
15	-2.0208	56	1.4911	15	-1.7374	56	1.7060	15	-1.7003	56	1.0167
16	-1.9249	57	1.6399	16	-1.6384	57	1.8496	16	-1.6126	57	1.0950
17	-1.8325	58	1.8043	17	-1.5433	58	2.0086	17	-1.5285	58	1.1763
18	-1.7432	59	1.9890	18	-1.4517	59	2.1881	18	-1.4477	59	1.2608
19	-1.6566	60	2.2017	19	-1.3630	60	2.3954	19	-1.3697	60	1.3490
20	-1.5723	61	2.4550	20	-1.2769	61	2.6432	20	-1.2942	61	1.4414
21	-1.4900	62	2.7725	21	-1.1932	62	2.9552	21	-1.2209	62	1.5388
22	-1.4095	63	3.2076	22	-1.1115	63	3.3847	22	-1.1496	63	1.6420
23	-1.3306	64	3.9299	23	-1.0316	64	4.1015	23	-1.0800	64	1.7520
24	-1.2529	65	N/A	24	-0.9533	65	N/A	24	-1.0120	65	1.8702
25	-1.1764			25	-0.8764			25	-0.9454	66	1.9983
26	-1.1009			26	-0.8007			26	-0.8799	67	2.1388
27	-1.0262			27	-0.7260			27	-0.8156	68	2.2949
28	-0.9522			28	-0.6522			28	-0.7522	69	2.4716
29	-0.8786			29	-0.5792			29	-0.6897	70	2.6764
30	-0.8055			30	-0.5068			30	-0.6279	71	2.9219
31	-0.7326			31	-0.4349			31	-0.5667	72	3.2319
32	-0.6598			32	-0.3633			32	-0.5060	73	3.6596
33	-0.5871			33	-0.2919			33	-0.4458	74	4.3748
34	-0.5142			34	-0.2206			34	-0.3860	75	N/A
35	-0.4411			35	-0.1492			35	-0.3264		
36	-0.3677			36	-0.0777			36	-0.2670		
37	-0.2937			37	-0.0059			37	-0.2077		
38	-0.2191			38	0.0663			38	-0.1485		
39	-0.1438			39	0.1391			39	-0.0892		
40	-0.0676			40	0.2126			40	-0.0297		

Table 3.A.2 2007 Raw to Theta Conversions for Mathematics Grades 2, 3, and 4

Grade 2				Grade 3				Grade 4			
Raw Score	Theta	Raw Score	Theta								
0	N/A	41	-0.3145	0	N/A	41	-0.2248	0	N/A	41	0.0696
1	-5.5492	42	-0.2363	1	-5.3782	42	-0.1480	1	-4.8964	42	0.1418
2	-4.8214	43	-0.1567	2	-4.6530	43	-0.0699	2	-4.1787	43	0.2152
3	-4.3816	44	-0.0757	3	-4.2158	44	0.0098	3	-3.7483	44	0.2902
4	-4.0599	45	0.0070	4	-3.8967	45	0.0913	4	-3.4357	45	0.3669
5	-3.8030	46	0.0917	5	-3.6423	46	0.1749	5	-3.1874	46	0.4457
6	-3.5871	47	0.1787	6	-3.4289	47	0.2608	6	-2.9798	47	0.5267
7	-3.3997	48	0.2683	7	-3.2439	48	0.3494	7	-2.8003	48	0.6103
8	-3.2329	49	0.3610	8	-3.0795	49	0.4412	8	-2.6413	49	0.6969
9	-3.0820	50	0.4571	9	-2.9310	50	0.5365	9	-2.4978	50	0.7871
10	-2.9436	51	0.5572	10	-2.7948	51	0.6360	10	-2.3667	51	0.8813
11	-2.8152	52	0.6620	11	-2.6687	52	0.7403	11	-2.2453	52	0.9802
12	-2.6951	53	0.7724	12	-2.5508	53	0.8503	12	-2.1321	53	1.0847
13	-2.5818	54	0.8893	13	-2.4397	54	0.9670	13	-2.0256	54	1.1957
14	-2.4744	55	1.0142	14	-2.3345	55	1.0918	14	-1.9248	55	1.3148
15	-2.3719	56	1.1487	15	-2.2341	56	1.2265	15	-1.8288	56	1.4436
16	-2.2737	57	1.2954	16	-2.1380	57	1.3735	16	-1.7371	57	1.5845
17	-2.1791	58	1.4575	17	-2.0455	58	1.5361	17	-1.6489	58	1.7408
18	-2.0878	59	1.6398	18	-1.9562	59	1.7192	18	-1.5638	59	1.9173
19	-1.9992	60	1.8500	19	-1.8697	60	1.9303	19	-1.4815	60	2.1219
20	-1.9131	61	2.1007	20	-1.7856	61	2.1824	20	-1.4016	61	2.3670
21	-1.8291	62	2.4156	21	-1.7036	62	2.4988	21	-1.3238	62	2.6762
22	-1.7470	63	2.8477	22	-1.6235	63	2.9330	22	-1.2478	63	3.1030
23	-1.6665	64	3.5674	23	-1.5449	64	3.6545	23	-1.1735	64	3.8164
24	-1.5875	65	N/A	24	-1.4678	65	N/A	24	-1.1005	65	N/A
25	-1.5096			25	-1.3919			25	-1.0287		
26	-1.4329			26	-1.3171			26	-0.9580		
27	-1.3570			27	-1.2431			27	-0.8882		
28	-1.2819			28	-1.1698			28	-0.8191		
29	-1.2073			29	-1.0972			29	-0.7506		
30	-1.1333			30	-1.0250			30	-0.6826		
31	-1.0596			31	-0.9531			31	-0.6150		
32	-0.9861			32	-0.8814			32	-0.5475		
33	-0.9127			33	-0.8098			33	-0.4802		
34	-0.8393			34	-0.7382			34	-0.4128		
35	-0.7658			35	-0.6664			35	-0.3454		
36	-0.6920			36	-0.5943			36	-0.2776		
37	-0.6178			37	-0.5218			37	-0.2095		
38	-0.5431			38	-0.4487			38	-0.1408		
39	-0.4677			39	-0.3750			39	-0.0715		
40	-0.3916			40	-0.3004			40	-0.0014		

Chapter 4: Validity

This chapter summarizes evidence supporting the content and convergent validity of the California Standards-based Tests in Spanish. The content validity evidence is based on the spring 2007 test assembly process. The convergent validity is based on a study relating the CSTs and the STS using the 2007 data.

Validity Evidence Based on Test Content

Content validity refers to the degree to which the content of a test is congruent with the purpose of testing, as determined by subject matter experts. STS items were developed to align with the content standards that are representative of the broader content domains: reading/language arts and mathematics. Thus, the content-related evidence of validity concerns the extent to which the test items represent these specified content domains and cognitive dimensions.

Content validity also provides information about how well an item measures its intended construct. Such validity is determined by a critical review of the items by experts in the field. For the STS, these reviews are conducted by experts in their designated areas from both the CDE and ETS. For these reviews, ETS senior content staff worked directly with CDE content consultants.

The CDE content consultants each have extensive experience in K–12 assessments, particularly in their subject of expertise, and many are former teachers. At minimum, each CDE content consultant holds a bachelor's degree; most have advanced degrees in their area of expertise. All ETS content and test development staff have extensive experience with K–12 assessments, experience in teaching students with a broad range of abilities, and an understanding of the California standards. They each hold, at minimum, bachelor's degrees; most have advanced degrees within their areas of expertise.

After the STS items had been written by ETS-trained bilingual/biliterate item writers in Spanish and English, a series of reviews, including reviews by ETS content assessment specialists and the external ARPs, were conducted to ensure that each item was measuring the appropriate California Content Standard and was matched to the item specifications. A description of the STS reporting clusters and the standards associated with each reporting cluster is provided in Appendix 2.B in Chapter 2. Detailed information on the item and content evaluation process can also be found in Chapter 2.

STS Assessment Review Panel

In addition to the thorough content reviews completed by ETS content-area experts and the content staff at the CDE, all STS items are reviewed by a content-area ARP. The majority of the ARP content-area reviewers are bilingual and biliterate in Spanish and English. The ARPs are advisory panels to ETS on areas related to item development for the STS. Their credentials are presented later in this chapter.

Purpose

As described in Chapter 2, ETS is responsible for working with ARPs as items are developed for the STS tests. For the 2007 development cycle, 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 item content, clarity of phrasing, and item quality. ETS provided the ARPs with the opportunity to review the items with the applicable field-test statistics and to make recommendations for the use of items in subsequent test forms. The ARPs may raise concerns in their examination of test items related to age/grade appropriateness and to gender, racial/ethnic, and socioeconomic bias.

Since the ARPs are responsible for reviewing the newly developed items for alignment to the California content standards, they determine if the items are:

- Measuring the California standards as appropriate for the STS testing population
- Free from bias
- Interesting and appropriate to students tested at a particular grade/course level

Composition

The ARPs are comprised of current and former teachers, resource specialists, administrators, curricular experts, and other education professionals. Current school staff members must meet 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 subject area related to reading/language arts or mathematics; and
- Knowledge and experience with the California content standards in reading/language arts or mathematics.

School administrators, district/county content/program specialists, or university educators serving on the STS ARPs must meet the following qualifications:

- Three or more years of experience as a school administrator, district/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 or mathematics; and
- Knowledge of and experience with the California content standards in reading/language arts or mathematics.

Every effort is made to ensure that ARP committees include representation of gender and of the geographic regions and ethnic groups in California. Efforts are also made to ensure representation by members with experience serving California's diverse special education population.

Current ARP members were recruited through an application process. Recommendations were solicited from districts and county offices of education in addition to CDE and SBE staff. Applications were received and reviewed throughout the year. ARP applications were reviewed by the ETS Assessment Directors and Test Development Project Lead, who confirmed that the applicant's qualifications met the specified criteria. Applications that met the criteria were forwarded to the CDE and SBE staff for review and final approval. Upon approval, the applicant was notified that he or she had been selected to serve on the ARP committee. Table 4.1 shows the educational qualifications, present occupation and credentials of the current STS ARP members.

Table 4.1 STS ARP Member Qualifications, by subject and total

	RLA	Math	Total
Total	22	19	41
Occupation (Members may teach multiple levels)			
Teacher or Program Specialist, Elementary/Middle School	17	7	24
Teacher or Program Specialist, High School	1	6	7
Teacher or Program Specialist, K–12	1	0	1
University Personnel	2	5	7
Other District Personnel (e.g., Director of Special Services, etc.)	0	1	1

	RLA	Math	Total
Highest Degree Earned			
Bachelor's Degree	22	19	41
Master's Degree	18	13	31
Doctorate	6	7	13
Credential (Members may hold multiple credentials)			
Elementary Teaching (Multiple Subjects)	13	4	17
Secondary Teaching (Single Subject)	3	6	9
Special Education	0	0	0
Reading Specialist	0	0	0
English Learner (CLAD, BCLAD)	12	2	14
Administrative	9	4	13
Other	2	2	4
None (teaching at university level)	2	5	7

Currently, there are no term limits for ARP members. While most members participate in the ARP meetings for only one STAR testing program, some members serve on more than one panel to encourage consistency in decisions among the STAR testing programs. ETS and CDE annually review the ARP membership for active participation. Members who have not attended a meeting within the last two years are notified that their invitation to participate may be withdrawn due to lack of attendance at meetings. In addition, ETS and CDE regularly review concerns about members whose conduct may be unprofessional and not conducive to the purpose of the ARP. If the concerns are determined to be valid, membership is revoked immediately.

STS Item Writers

The items selected for each STS test are written by special panels of item writers with expertise in 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. Thus, the participants were particularly experienced in writing to the standards assessed on STS. All item writers met the following minimum qualifications:

- The minimum 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.
- At least three years of classroom teaching experience at the appropriate grade level
- 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
- Bilingual and biliterate in Spanish and English

Participants attended a general STS item development training session, and then were given specific subject-area training. After viewing multiple examples of previously written STS items, participants were given item writing assignments. ETS facilitators provided feedback, and peer review methods were employed.

Additional information about the item writing process is described in Chapter 2.

STS Development Procedures

The STS exams were constructed to measure the California content standards as well as to meet psychometric criteria for difficulty and reliability. The psychometric criteria were evaluated using projections based on item statistics from field-testing or previous operational administrations.

Test Assembly

Test blueprints for the components of the STAR Program (which includes STS) were proposed by ETS, reviewed and approved by the respective ARPs, also reviewed and approved by the CDE, and was adopted by the SBE in July 2005. For each test, the California content standards were used as the basis for choosing items. Additional technical targets (e.g., difficulty and discrimination) for test construction were established with the goal of maintaining parallel forms across years to the greatest extent possible.

Test Specifications

Statistical Specifications

The primary statistical targets used for STS test assembly in 2007 were the test information functions based on the IRT item parameters (one-parameter model) and average point biserial. When Rasch IRT estimates are used, the target information function makes it possible to choose items to produce a test that has the desired level of difficulty. The point biserial correlation is a measure of how well the items discriminate among test-takers and is related to the overall reliability of the test.

These specifications were developed from the analyses of field-test items administered in the fall of 2006; the target values and ranges for the specifications are presented in Table 2.1 on page 6. The minimum target value for a proportion-correct value (indicator of item difficulty) was set at 0.20 and maximum was set at 0.95; the minimum item point biserial was set at 0.14 for each test. The target mean and standard deviation of item b-values consistent with the information curves were also provided to test development staff to help with the test construction process.

Content Specifications

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

Item Development

ETS senior content staff led the item writers in the item development and review process. In addition, experienced ETS content specialists and assessment editors reviewed each item during the forms construction process. The lead assessment specialist for each content area worked directly with the other ETS assessment specialists to carefully review and edit each item for technical characteristics like quality (for example, one right answer, clearly stated stem, absence of clueing, plausibility of, distractors), match to standard, and conformity with California-approved item-writing practices.

ETS has maintained item specifications for each STS test. ETS followed the SBE-approved Item Utilization Plan to guide the development of the quantity of items for each subject area.

Item specification documents included the constructs to be measured and the California content standards included in the test blueprints. The item specifications help ensure that the STS tests consistently match the content standards from year to year. Item writing emphasis is determined in consultation with the CDE. The item specifications also provide specific and important guidance to item writers, and ensure that items are consistent in approach and written to measure the standards. The item specifications describe the general characteristics of the items within each content standard, indicate item types or content to be avoided, and define the content limits for the items. In summary, the specifications included the following:

- A statement of the strand or topic for the standard
- A full statement of the academic content standard, as found in each STS blueprint
- The expected cognitive level(s) of items written for the standard (Acquire, Integrate, Extend), as defined by ETS and approved by CDE
- The construct(s) appropriately measured by the standard
 - A description of the kinds of stems appropriate for multiple-choice items for the standard
 - A description of the kinds of distracters appropriate for multiple-choice items for the standard
 - A description of specific kinds of items to be avoided, if any (e.g., no RLA items about insignificant details)
 - A description of appropriate stimuli (e.g., charts, tables, graphs, or other artwork) for mathematics items
 - The content limits for the standard (e.g., one or two variables, maximum place values of numbers) for mathematics items
- A description of appropriate reading passages (if applicable) for RLA items

Internal Reviews

After the items were written, ETS employed a series of internal reviews. The reviews established the criteria used to judge the content validity of the items, making sure that each item was measuring 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 CDE and the ARPs. Because of the complexities involved in producing defensible items for high-stakes programs such as the STAR Program, it is essential that many experienced individuals review each item before it is brought to CDE and the ARP and SPAR panels.

The ETS review process for the STS included the following:

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

Throughout this multi-step review process, the lead content area assessment specialists and development team members continually evaluated the relevance of the information being assessed, its relevance to the California content standards, its match to the test and item specifications, and its appropriateness to the population being assessed. Items that are only peripherally related to the test and item specifications, that do not measure core outcomes reflected in the California content standards, or that are not developmentally appropriate were eliminated early in this rigorous review process.

1. Internal Content Review

STS items and materials received two reviews from the content area assessment specialists. These assessment specialists made sure that the items and related materials were in compliance with ETS's written guidelines for clarity, style, accuracy, and appropriateness for California students, and in compliance with the approved item specifications. Assessment specialists reviewed each item following the criteria below:

- Relevance of each item as the item relates to the purpose of the test
- Match of each item to the item specifications, including cognitive level
- Match of each item to the principles of quality item development
- Match of each item to the identified standard
- 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 artwork, graphs, figures, etc.

The assessment specialists also checked all items against their cluster classification codes, both to evaluate the correctness of the classification and to ensure that a given task is of a type appropriate to the outcome it was intended to measure. The reviewers accepted the item and classification as written, suggested revisions, or recommended that the items be discarded. These steps occurred prior to CDE review.

2. 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 it was prepared for CDE and ARP review. In addition, 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 groups, individuals with disabilities, and women) to the history and culture of the United States and the achievements of individuals within these groups

3. Internal Editorial Review

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

ARP Meetings for Review of STS Items

The 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 steps:

- An overview of the purpose and scope of the STS
- An overview of STS test design specifications and blueprints
- An analysis of STS item specifications
- Review and evaluation of items for bias and sensitivity issues

The criteria for evaluating test items included:

- Overall technical quality
- Match to the California content standards
- Match to the construct being assessed by the standard
- Difficulty range
- Clarity
- Correctness of the key responses
- Plausibility of the distracters
- Bias and sensitivity factors

The committee was also trained on how to make recommendations for revising items. Guidelines for reviewing items were provided by ETS and approved by CDE. The guidelines for reviewing items are summarized below:

Item Guidelines: Does the item...

- Measure the content standard?
- Match the test item specifications?
- Align with the construct being measured?
- Test worthwhile concepts or information?
- Include administrator directions that give the student a full sense of what the item is asking?
- Avoid unnecessary wordiness?
- Reflect content that is free from bias against any person or group?

Stimulus Guidelines: 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, panel 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. These recommendations were recorded in a master item review booklet. Item review binders and other item evaluation materials also identified potential bias and sensitivity factors the ARP considers as part of its item reviews.

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.

SPAR Panel

The SPAR panel is responsible for reviewing and approving a single achievement test to be used statewide for the testing of Spanish-speaking English learners in California public schools, grades 2–4. At the SPAR panel meetings, all new items were presented in binders for review. The SPAR panel representatives ensured that the test items conformed to the requirements of *Education Code* Section 60614. If the SPAR panel rejected specific items, the items were replaced. For the SPAR panel meeting, the item development coordinator or an ETS content specialist who had been requested in advance by CDE attended the opening session and remained at a nearby location or near a telephone to be available to respond to any questions during the course of the meeting.

Validity Evidence Based on Relations to Other Variables

Analyses of the relationship of test scores to variables external to the test provide an important source of validity evidence, which address questions about the degree to which these relationships are consistent with the construct underlying the proposed test interpretations. Relationships between test scores and other measures intended to measure similar constructs provide convergent validity evidence. For STS reading/language arts and mathematics, the convergent evidence can be collected through examining the relationship between the STS tests and their CST counterparts.

CSTs assess students in English–language arts, mathematics, history–social science, and science. All students who take the STS are also required to take the CSTs at their grade level. CST and STS tests measure the same California content standards except that they are given in different languages. The STS math tests are expected to relate closely to the CST math tests as they are intended to measure the same construct of math ability. The STS RLA tests should relate somewhat to the CST

ELA, because both measure the construct of reading ability. However, given that STS RLA measures reading in Spanish and the CST ELA measures reading in English, the correlation will be limited. Two groups of correlations were examined:

- Correlations between STS math tests and CST math tests; and
- Correlations between STS RLA tests and CST ELA tests.

Table 4.A.1 and Table 4.A.2 on page 39 present correlations between 2007 CST and STS test scores based on the matched observations of students who obtained valid scores on both STS and CST. Over 97% of the STS examinees were able to be matched to their CST records. Most of the cases were matched based on the unique student ID. A few were matched based on student name and 14-digit school code.

Correlations were computed based on both the overall STS population and the STS target population. For each correlation index, the total STS sample size before matching, and the matched sample size, based on which correlation was computed, are provided immediately before the correlation in the tables. The results showed that STS math scores correlated highly with CST test scores, with correlations consistently around .85 across all grades. The STS RLA exhibited a moderately high correlation with the CST ELA tests for the corresponding grades, with correlations ranging from 0.61 to 0.68. This result was expected, as the CST ELA and STS RLA were designed to measure reading in different languages, and students were not expected to be equally proficient in the two languages.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. *Standards for Educational And Psychological Testing*. Washington, D.C.: American Psychological Association, 1999.

Sireci, S. G. "Validity," *Encyclopedia of Psychological Assessment*. London: Sage Publications, 2002.

Wainer, H. and Braun, H. I. (1998). *Test Validity*. Mahwah, NJ: Lawrence Erlbaum Associates, 1998.

Appendix 4.A—Test Correlations Tables

Table 4.A.1 Correlations Between STS Mathematics Tests and CST Mathematics Tests

Grades	All STS Population			Target STS Population		
	Total No. of Examinees	No. of Examinees Matched	Math (STS)/ Math (CST)	Total No. of Examinees	No. of Examinees Matched	Math (STS)/ Math (CST)
2	18,665	18,369	0.85	12,469	12,348	0.85
3	13,031	12,772	0.85	8,926	8,842	0.85
4	7,943	7,744	0.85	5,283	5,223	0.85

Table 4.A.2 Correlations Between STS RLA Tests and CST ELA Tests

Grades	All STS Population			Target STS Population		
	Total No. of Examinees	No. of Examinees Matched	RLA (STS)/ ELA (CST)	Total No. of Examinees	No. of Examinees Matched	RLA (STS)/ ELA (CST)
2	18,767	18,433	0.66	12,486	12,347	0.66
3	13,084	12,785	0.67	8,952	8,855	0.68
4	7,972	7,717	0.61	5,291	5,214	0.62

Chapter 5: Score Reports

This chapter describes the score reporting procedures and score-level analyses obtained for the spring 2007 administration of the Standards-based Tests in Spanish. The first section of the chapter describes the score scales, followed by a discussion on the types and purposes of score reports that are relevant to STS and the related STAR components. The second section focuses on summaries of scores obtained on the spring 2007 administration of the STS. Following the pattern of previous chapters, the results are reported for each subject area and for each grade within a subject area. The numbers in the summary tables may not match exactly to the results reported on the CDE Web site, as there may be slight differences in the samples used to compute the statistics.

Descriptions of Scores

Raw Score

For all of the tests the raw score is simply the sum of correct responses on the test items.

Percent Correct Score

Percent correct scores are raw scores divided by the total number of items on the test.

Score Reporting

The tests that make up the STAR Program provide results or score summaries that are reported for different purposes. The three major purposes include:

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 STAR reports is presented in the next section.

Score Report Applications

STAR Program results provide parents and guardians with information about their children's progress. The results are a tool for increasing communication and collaboration between parents, guardians, and teachers. Along with teacher report cards and information from school and classroom tests, the STAR STS Student Report can be used by parents and guardians to talk with teachers about ways to improve their children's achievement of the California content standards. Any discrepancies between performance reported on report cards and the scores reported on the STAR STS Student Report should also be discussed.

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

STAR Program results help school districts and schools identify strengths and weaknesses in their instructional programs. Each year, school districts and school staffs examine STAR Program test results at each grade level and subject tested. Their findings are used to help determine:

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

Score Report Contents

The individual STAR Student Report provides overall percent correct, the number correct, and the percent correct in each reporting cluster (subscore) for each STS taken by the student. The overall percent correct indicates how well a student is achieving the California content standards tested. In

addition, STS results are reported in clusters that describe how a student performs on a set of state content standards. Reporting cluster results compare an individual student's percent correct score to the average percent correct for the state, as a whole, and for the district where a student attends school.

Reports for students with disabilities, who use accommodations or modifications, include a notation about the test administration that indicates:

- Student used accommodations, or
- Student was tested with modifications.

Scores for students who use accommodations are reported in the same way as for non-accommodated tests.

In addition to individual student reports, several other reports are also provided to different groups of stakeholders. A description of those reports is provided in Appendix 5.A.

Score Distributions and Summary Statistics

This section summarizes and reports the score level analyses including descriptive statistics on raw and percent correct scores, and the distribution of examinees in various raw score ranges. The analyses were performed on the sets of valid scores for both the overall population and the target population on each STS. Just as all the score reports are based only on the operational items (i.e., scores on field-test items are not included), the statistics reported in this chapter are based solely on the operational items.

The descriptive information including number of items on each STS, number of examinees taking the STS, and the corresponding means and standard deviations of raw and percent correct scores for each STS are presented in Table 5.1 and Table 5.2 for the overall and target population respectively. The average raw scores for the target population were slightly but consistently higher than those for the overall population.

Table 5.1 Mean and Standard Deviation of Raw and Percent Correct Scores for STS Overall Population

Subject	STS	No. of Items	Overall Population				
			No. of Examinees	Raw Score		Percent Correct Score	
				Mean	Std. Dev.	Mean	Std. Dev.
Reading/Language Arts	2	65	18,767	40.49	12.20	62.30	18.77
	3	65	13,084	36.67	11.51	56.41	17.71
	4	75	7,972	40.20	13.58	53.60	18.11
Mathematics	2	65	18,665	44.45	11.52	68.38	17.72
	3	65	13,031	43.14	12.76	66.36	19.63
	4	65	7,943	39.25	13.14	60.38	20.21

Table 5.2 Mean and Standard Deviation of Raw and Percent Correct Scores for STS Target Population

Subject	STS	No. of Items	Target Population				
			No. of Examinees	Raw Score		Percent Correct Score	
				Mean	Std. Dev.	Mean	Std. Dev.
Reading/Language Arts	2	65	12,486	40.82	12.05	62.80	18.54
	3	65	8,952	37.07	11.34	57.03	17.44
	4	75	5,291	40.47	13.55	53.96	18.07
Mathematics	2	65	12,469	44.84	11.33	68.99	17.43
	3	65	8,926	43.78	12.46	67.35	19.17
	4	65	5,283	39.93	12.90	61.44	19.84

Table 5.B.1 and 5.B.2 in Appendix 5.B on page 45 show the distributions of raw scores by STS for the overall and target population respectively. The tables show the distribution of examinees within each 5 point raw score interval for the 6 STS tests. The raw scores range from 0 to 75 for RLA grade 4, resulting in 15 score intervals. For all other STS tests, raw scores range from 0 to 65, resulting in 13 score intervals.

Reference

California Department of Education, “2007 STAR CST-CAT/6 Survey, CAPA, and STS Printed Reports,” <http://www.startest.org/pdfs/STAR.reports.2007.pdf>, 2007.

Appendix 5.A—Types of Score Reports Tables

Table 5.A.1 Score Reports Reflecting STS Results

2007 STAR STS REPORTS	
DESCRIPTION	DISTRIBUTION
The STAR STS Student Report	
<p>This report provides parents/guardians and teachers with the student’s results in tables and graphs. It mainly includes:</p> <ul style="list-style-type: none"> • Overall percent correct for each content area • Number and percent correct in each reporting cluster for each content area 	<p>This report includes individual student results and is not distributed beyond parents/guardians and the student’s school.</p> <p>Two color copies of this report are provided for each student: One is for the student’s current teacher, and one is to be distributed to parents/guardians by the district.</p>
Student Record Label STS	
<p>These reports are printed on adhesive labels to be affixed to the student’s permanent school records. Significant information includes:</p> <ul style="list-style-type: none"> • Overall percent correct for each content area 	<p>This report includes individual student results and is not distributed beyond the student’s school.</p>
Student Master List STS	
<p>This report is an alphabetical roster of individual student results. It mainly includes:</p> <ul style="list-style-type: none"> • Overall percent correct for each content area • Number and percent correct in each reporting cluster for each content area 	<p>This report provides administrators and teachers with a quick reference to all students’ results within each grade or within each grade and year-round schedule at a school.</p> <p>This report includes individual student results and is not distributed beyond the student’s school.</p>
Student Master List Summary STS	
<p>This report summarizes student results at the school, district, county, and state level for each grade. It does <i>not</i> include any individual student information. The following data is summarized by subject:</p> <ul style="list-style-type: none"> • Number of students enrolled, number and percent of students tested, and number and percent of valid scores • Mean percent correct, and standard deviation for each subject area tested • The number of items and the mean percent correct for each reporting cluster 	<p>This report is a resource for evaluators, researchers, teachers, parents/guardians, community members, and administrators.</p> <p>One copy is sent to the school and one to the district. This report is also produced for districts, counties, and the state.</p> <p>Note: The data on this report may be shared with parents/guardians, community members, and the media only if the data are for 11 or more students.</p>

2007 STAR STS REPORTS	
DESCRIPTION	DISTRIBUTION
Subgroup Summary STS	
<p>This set of reports disaggregates and reports results by the following subgroups:</p> <ul style="list-style-type: none"> • All students • Disability Status • Economic status • Gender • English-language fluency • Primary Ethnicity <p>These reports contain no individual student-identifying information and are aggregated at the school, district, county, and state level.</p> <p>For each subgroup within a report, and for the total number of students, the following is included:</p> <ul style="list-style-type: none"> • Total number tested in the subgroup • Percent tested in subgroup as a percent of all students tested • Number and percent of valid scores • Number tested who received scores 	<p>This report is a resource for evaluators, researchers, teachers, parents/guardians, community members, and administrators.</p> <p>One copy is sent to the school and one copy to the district. This report is also produced for districts, counties, and the state.</p> <p>Note: The data on this report may be shared with parents/guardians, community members, and the media only if the data are for 11 or more students.</p>

Appendix 5.B—Raw Score Distribution Tables

Table 5.B.1 Distribution of Raw Scores for STS Overall Population

Raw Score	Reading/Language Arts			Mathematics		
	Grade 2	Grade 3	Grade 4	Grade 2	Grade 3	Grade 4
71 – 75	N/A	N/A	4	N/A	N/A	N/A
66 – 70	N/A	N/A	92	N/A	N/A	N/A
61 – 65	240	83	416	1,209	916	346
56 – 60	1,448	614	746	2,501	1,702	730
51 – 55	2,877	1,085	867	2,772	1,882	882
46 – 50	3,084	1,492	956	2,883	1,653	831
41 – 45	2,729	1,810	913	2,674	1,570	913
36 – 40	2,237	1,806	884	2,195	1,479	922
31 – 35	1,808	1,870	831	1,904	1,284	946
26 – 30	1,533	1,800	848	1,271	1,113	945
21 – 25	1,330	1,429	801	815	841	800
16 – 20	1,064	852	479	358	462	473
11 – 15	385	232	126	79	120	142
06 – 10	32	11	9	3	9	13
00 – 05	0	0	0	1	0	0

Table 5.B.2 Distribution of Raw Scores for STS Target Population

Raw Score	Reading/Language Arts			Mathematics		
	Grade 2	Grade 3	Grade 4	Grade 2	Grade 3	Grade 4
71 – 75	N/A	N/A	1	N/A	N/A	N/A
66 – 70	N/A	N/A	63	N/A	N/A	N/A
61 – 65	162	64	286	822	630	236
56 – 60	972	420	512	1,695	1,224	515
51 – 55	1,968	762	584	1,905	1,348	589
46 – 50	2,106	1,061	633	1,964	1,179	586
41 – 45	1,847	1,264	607	1,881	1,101	652
36 – 40	1,471	1,276	586	1,460	1,037	653
31 – 35	1,177	1,303	574	1,198	824	615
26 – 30	1,029	1,204	546	776	727	603
21 – 25	844	937	502	498	521	468
16 – 20	655	524	312	221	261	278
11 – 15	237	131	79	45	69	80
06 – 10	18	6	6	3	5	8
00 – 05	0	0	0	1	0	0

Chapter 6: Test Fairness

In order to ensure equity among various subpopulations, comprehensive analyses were conducted for the spring 2007 administration of the Standards-based Tests in Spanish. This chapter summarizes the subgroup analyses performed at the test level. Detailed item level analyses were also conducted when sufficient sample sizes were available for a subgroup. In addition, analyses are presented related to students with physical and learning disabilities who took the test under standard or modified conditions.

The chapter is, therefore, divided into two major sections. The first section presents the summary of statistics obtained on various demographic indicators, while the second section discusses the distributions of examinees grouped by accommodation provisions. Following the pattern of previous chapters, all analyses are replicated for each grade within a subject area.

Since assuring test security is crucial in the sustenance of a fair test, the chapter also briefly describes procedures for ensuring test security.

Demographic Distributions

Table 6.1 presents a listing of various subgroups included in this chapter, along with their definitions. Summary statistics for all students, and for subgroups based on demographic variables presented in Table 6.1, are discussed in this section. The demographic variables examined included gender, country of origin, economic status, enrollment in U.S. schools, English learner program participation, and special education programs.

The results of the demographic-based analyses are presented in Tables 6.A.1 through 6.A.12 for the six STS tests. Two summary tables are provided for each STS test based on the overall and target population respectively. The tables include number of students tested for whom valid scores were available, mean number correct raw scores, and standard deviation of number correct raw scores, as well as mean percent correct scores within each reporting cluster. The statistics in these tables were based on all valid scores in the overall population. For demographic groups of fewer than 11 examinees, no summary statistics on the number correct raw score or cluster scores were presented.

Table 6.1 Subgroup Definitions

Subgroup	Definition
Gender	<ul style="list-style-type: none"> • Female • Male
Country of Origin	<ul style="list-style-type: none"> • United States • Mexico • Spain • Puerto Rico • Cuba • Guatemala • El Salvador • Columbia • Brazil • Ecuador • Venezuela • Peru • Bolivia • Chile • Paraguay • Argentina

Subgroup	Definition
	<ul style="list-style-type: none"> • Uruguay • Panama • Costa Rica • Other
Economic Status	<ul style="list-style-type: none"> • Economically Disadvantaged (NSLP) • Non-Economically Disadvantaged
Enrollment in U.S. Schools	<ul style="list-style-type: none"> • Less than 12 months • 12 months or more
EL Program Participation	<ul style="list-style-type: none"> • English Learner (EL) in English Language Development (ELD) • EL in ELD and specially designed academic instruction in English (SDAIE) • EL in ELD and SDAIE with Primary Language Support • EL in ELD and Academic Subjects through Primary Language • Other EL services • None (EL only)
Special Education Services	<ul style="list-style-type: none"> • Special Education Services • No Special Education Services

Target students constitute about two thirds of the overall STS population. They perform slightly better than the nontarget/optional students as shown in the table that the mean number correct raw score is 1 point higher for the target population than for the overall population for most of the STS tests.

The number of males testing was slightly higher than females for the grades 2 and 3 RLA and math tests. On the other hand, more females than males took the STS for Grade 4 RLA and Math tests. Females scored consistently higher than males on most STS tests. The differences between mean number correct raw scores were especially large for RLA tests, where females scored 3 to 5 points higher than males. The performance of males and females was comparable for grades 2 and 3 math. On the STS for Grade 4 Math, females again scored higher than males, by 2 points.

Country of origin is an optional field and most students chose not to fill in this field. Among students who did provide this information, most students coded *United States* or *Mexico* as their countries of origin.

The vast majority of STS students, approximately 90%, were classified as economically disadvantaged using the National School Lunch Program (NSLP) proxy. The proportion of students in the NSLP decreased slightly, by about 1%, as the grades increased. There was no noticeable pattern of performance difference between examinees who participated in NSLP and examinees who did not enroll for NSLP.

The percentages of STS students who have been in U.S. schools less than 12 months were approximately 15, 20, and 30 percent for grades 2, 3 and 4 respectively. The percentages were very similar for overall and the target STS population. Across all the STS tests, the students who have been in U.S. schools more than 12 months were found to perform considerably better than the students who have been in U.S. schools less than 12 months.

The summary statistics of the EL program participation showed that the majority of the STS takers participated in the ELD and Academic Subjects through Primary Language program. The proportion of students participating in this program was higher for the target population, which ranged from .74 to .88, than for the overall population, which ranged from .55 to .69, because of the way the target population is defined (that is, Spanish-speaking English learners need to participate in this program or be in U.S. schools less than 12 cumulative months to be counted as a target STS examinee). The participation rates for this program decreased as the grade increased. Generally, the students who participated in this EL program were found to outperform the students who participated in other EL

programs, but it is not possible to know whether this difference is due to differences in the programs, selection of students for the programs, or other correlated variables.

The approximate percentage of overall STS population who received special education services ranged from 3 to 5% for all STS tests. A slightly higher proportion of the target STS population, between 4 to 6%, received special education services.

Test Variations, Accommodations and Modifications

STS test-takers include students with disabilities. Most students with disabilities take the STS tests under standard conditions. Some students with disabilities, however, may need assistance when taking the STS tests. This assistance takes the form of test variations, accommodations, or modifications. All students in these categories may have test administration directions simplified or clarified. In addition, all eligible students may have test variations if they are regularly used in the classroom. They also must be allowed to use the accommodations and modifications that are specified in each student's individualized education program (IEP) plan or Section 504 Plan. These accommodations and/or modifications must match the one(s) used for classroom work throughout the year.

Test variations, accommodations, and modifications for the statewide assessments, including the STS Program, are defined in the next sections.

Category 1: Test Variations

Eligible students may have test variations if regularly used in the classroom. For example, students may be tested in a smaller group or individually, have special lighting or adaptive furniture, or use magnifying equipment.

Category 2: Accommodations

Eligible students are permitted to take the STS with accommodations if specified in the student's IEP plan or Section 504 Plan for use on the STS or for use during classroom instruction and assessment. Examples of accommodations are large-print or braille versions of the STS or providing more than one day for a test designed for a single sitting.

Category 3: Modifications

Eligible students are permitted to take the STS with modifications if specified in the student's IEP plan or Section 504 Plan for use on the STS or for use during classroom instruction and assessment. Examples of modifications include an examiner's reading the test to the student or a student's using a calculator to perform computations on the mathematics test.

Appendix 6.B presents the Matrix of Test Variations, Accommodations, and Modifications for Administration of California Standards-based Tests in Spanish. The matrix provides a complete list of the variations, accommodations, and modifications that were allowed under the STAR STS Program in 2007.

Accommodations change the way the test is given but do not change what is tested. Modifications fundamentally change what is being tested. The purpose of test variations, accommodations, and modifications is to enable the students to take the STS, not to give them an advantage over other students or to improve their scores.

Accommodation Summaries

The percentage of students utilizing various testing accommodations during the 2007 administration of all six STS tests is presented in Appendix 6.C. The data is organized into five sections within each table. The first section presents the percentages of students for the total testing population. The second section presents the results for target and nontarget STS takers. The third section presents the results for Special Education and non-Special Education students. The fourth

section presents the results for students who are in U.S. schools less than 12 months and students who are in U.S. schools more than 12 months. The final section presents the results for various categories based on EL program participation. Most accommodations are common across different STS tests. Additional accommodations were included for the STS for math comprising of the use of calculators, arithmetic tables, and math manipulatives.

Examinees who utilized various accommodations, modifications, and variations constituted less than 1% of the overall examinee population. As expected, a significantly large percentage of Special Education students made use of the accommodations regardless of the grade and STS administered as compared to the non-Special Education students. Also, students in U.S. schools more than 12 months used these services more frequently than students in U.S. schools less than 12 months.

The most frequently used modifications or accommodations were the use of supervised breaks, having examiners read the questions aloud, and administration of the test at the most beneficial time of day to the student for both RLA and math. Students at the lower grade levels also frequently used the accommodation of being tested over more than one day.

The use of any modification, accommodation, or variation was less frequent for RLA than for math. For math, there were considerably more students who had the examiners read the questions aloud than for RLA.

Of all students making use of the accommodations or modifications, the largest percentage used IEP accommodations or modifications.

DIF Analyses

Differential Item Functioning (DIF) analyses measure differences in item performance between different demographic groups of students who have similar overall test performance.

DIF analyses were performed on all operational items and all field-test items for which sufficient student samples were available. The sample size requirements for the field-test 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 DIF statistic (Mantel and Haenszel, 1959; Holland and Thayer, 1985). This statistic is based on the estimate of constant odds ratio and is described as:

$$\alpha_{MH} = \frac{\left(\sum_m R_{rm} \frac{W_{fm}}{N_{tm}} \right)}{\left(\sum_m R_{fm} \frac{W_{rm}}{N_{tm}} \right)} \quad (6.1)$$

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

$$MH \ D - DIF = -2.35 \ln [\alpha_{MH}] \quad (6.2)$$

where,

- R = number right,
- W = number wrong,
- N = total in:
- f_m = focal group at ability level m,
- r_m = reference group at ability level m, and
- t_m = total group at ability level 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 DIF values. These categories have been used by all ETS testing programs for more than 13 years. The definitions of the categories based on evaluations of the item-level MH D-DIF statistics is as follows:

DIF Category	Definition
A (negligible)	MH D-DIF not significantly different from zero, or has an absolute value less than one.
B (moderate)	MH D-DIF is significantly different from zero, and is either (1) less than 1.5; or (2) not significantly different from one.
C (large)	MH D-DIF is significantly different from one, and has an absolute value greater than 1.5.

The groups studied for DIF were based on gender; the sample sizes for other groups were too small to conduct dependable analyses. The minimum required sample sizes for conducting DIF analyses were 100 examinees in the reference as well as the focal group and 400 examinees combining reference and focal group. The results of the DIF analyses are presented in Appendix 6.D. In these tables, classifications of A-, B-, or C- indicate DIF against the focal group (for example, the female group) and classifications of A+, B+, and C+ indicate DIF in favor of the focal group. There were no operational items that were identified as exhibiting significant DIF. Table 6.D.1 represents the field-test items exhibiting significant DIF. There is only one item that is classified as a male–female DIF item. Test developers have been instructed to avoid selecting field-test items flagged as having shown DIF that disadvantage a focal group (C-DIF) for future operational test forms unless their inclusion is deemed essential to meeting test-content specifications.

Table 6.D.2 summarizes the DIF category classifications for operational items in each STS test based on the A, B, or C classifications. Table 6.D.3 summarizes the DIF category classifications for the field-test items. Both tables are presented for all six STS tests. Most operational and field-test items are classified into A- and A+ categories. For grade 4 RLA and math, DIF analyses were not conducted for the majority of the field-test items because of small sample sizes.

Test Security and Confidentiality

All tests within the STAR Program are secure documents. For the 2007 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), 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 student data. To ensure security for all the tests that ETS develops or handles, ETS maintains an Office of Testing Integrity (OTI), as discussed next.

ETS’s Office of Testing Integrity (OTI)

The OTI is a division of ETS that provides Quality Assurance and resides in the ETS Legal Department. The Quality Assurance division publishes and maintains *ETS Standards for Quality and Fairness*, which supports 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.

OTI’s mission is to

- Prevent and minimize any testing security violations that can impact the fairness of testing
- Prevent and investigate any security breach

- Report on security activities

OTI helps prevent misconduct on the part of test-takers and administrators, detect potential misconduct through empirically established indicators, and resolve situations in a fair and balanced way that reflects the laws and professional standards governing the integrity of testing.

Test Development

During the test development process, ETS staff members consistently follow these established security procedures:

- Only authorized individuals have access to test content at any step in the development, review, and data analysis processes.
- Test developers keep all hardcopy 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 development process.
- Test developers take further security measures whenever they share items outside of ETS, including registered, secure mail, express delivery, and tracking records of sending and receipt of any test materials.

Item Review

ETS enforces security measures at ARP meetings to protect the integrity of meeting materials using these guidelines:

- Individuals who participate in the ARPs must sign the confidentiality agreement.
- Meeting materials are strictly managed before, during, and after the review meetings.
- Meeting participants are supervised at all times during the meetings.
- Use of electronic devices is strictly prohibited in the meeting rooms.

Item Bank

Once the ARP review is complete, the items are placed in the item bank along with their corresponding review information. ETS then delivers the items to the CDE via a delivery of the STAR electronic item bank. Subsequent updates to items are based on field-test and operational use of the items. However, only the latest version of the item is in the bank at any point in time, along with the administration data from every administration that has 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 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 authorize access to the database or designated portions of the database. In addition, only users authorized to access the specific SQL database will be able to use the electronic item banking system. A designated administrator at the CDE and at ETS will authorize users.

Transfer of Forms and Items to CDE

ETS shares a File Transfer Protocol (FTP) site with the CDE. FTP is a standard method for exclusive routing of files. It is a password-protected server that only authorized users may access. On that site, ETS posts Word, PDF, or other document files for the CDE to review. ETS sends an e-mail to the CDE to notify them that files are posted. Item data are always transmitted in an encrypted format to the FTP site, never via e-mail.

Firewall

A firewall is software that prevents entry to files, e-mail, and other organization-specific programs from unauthorized users or computers. All ETS data exchange and internal e-mails remain within the ETS firewall at all ETS locations, from Princeton to San Antonio to Sacramento. The CDE has and will continue to view and approve ETS-developed application such as those on the STAR Management System at ETS's Sacramento office because the applications remain behind ETS's firewall before release. No hacker has ever broken into ETS's firewall.

Printing

After items and test forms are approved, the files are sent for printing on a CD using a secure courier system, such as Federal Express. According to established procedures, OTI pre-approves all printing vendors before they can work on secured confidential and proprietary test material. The printing vendor must submit a completed ETS Printing Plan and Typesetting Facility Security Plan which documents security procedures, access to test materials, work in progress, personnel procedures, and access to the facilities by the employees and visitors. After reviewing the completed plan, members of the OTI visit the printing vendor to conduct an onsite inspection. The secured printing vendor packs and ships printed test booklets to Pearson Educational Measurement (PEM) for packaging and distribution in a tight and precise way to prevent boxes from opening.

Test Administration

PEM receives testing materials from printers, packages them, and sends them to districts. After testing, districts return materials to PEM for scoring. During all these stages, PEM takes extraordinary measures to protect testing materials. PEM's customized Oracle business applications verify that inventory controls are in place from materials receipt to packaging. The reputable carriers used by PEM provide specialized handling and delivery service that maintain test security and meet the STS program schedule. The carriers provide inside delivery directly to the district STAR 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 district STAR coordinators are, therefore, required to keep all test materials in central locked storage except during actual test administration times. Test site coordinators are responsible for accounting for and returning all secure materials to the district coordinator, who is responsible for returning them to the STAR Scoring and Processing Centers. More specifically:

- District STAR coordinators have to sign and submit a "STAR Test (including field tests) Security Agreement for District and Test Site Coordinators" form to the STAR Technical Assistance Center before ETS can ship any testing materials to the district.
- Test site coordinators have to sign and submit a "STAR Test (including field tests) Security Agreement for District and Test Site Coordinators" form to the district STAR coordinator before any testing materials can be delivered to the school/test site.
- Anyone having access to the test materials sign and submit a "STAR Test (including field tests) Security Affidavit for Test Examiners, Proctors, Scribes, and any Other Person Having Access to STAR Tests" form to the test site coordinator before receiving access to any testing materials.

- It is the responsibility of each person participating in the STAR Program to report immediately any violation or suspected violation of test security or confidentiality. The test site coordinator is responsible for immediately reporting any security violation to the district STAR coordinator. The district STAR coordinator is to contact the CDE immediately and will be asked to follow up with a written explanation of the violation or suspected violation.
- Any irregularities in test security may result in invalidation of student test results.

Processing and Scoring

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

- There is controlled access to the facility.
- No test 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 Educational Measurement identification badges at all times in PEM facilities.
- No recording or photographic equipment is allowed in the scoring area without the consent of the CDE.

The completed and scored answer documents are then stored in secure warehouses. The only time they are touched then is if there is a dispute of a score. For example, districts or parents may request rescoring of a student's test. In such a case, a grade two or three test booklet or grade 4 answer document is removed from storage, copied, and sent securely to the ETS facility in Concord for handscoring, after which the copy is destroyed. No school or district personnel are allowed to look at the completed answer documents unless necessary for the purposes of transcription or to investigate irregular cases.

All answer documents and test booklets are destroyed after October 31 each year.

Transfer of Scores via Secure Data Exchange

After scoring is completed, PEM sends files to ETS and follows secure data exchange procedures. PEM provides overall security for assessment materials through its limited-access facilities and through its secure data processing capabilities. PEM enforces stringent procedures to prevent unauthorized attempts to access their facilities. Entrances are monitored by security personnel and a computerized badge-reading system is utilized. Upon entering the facilities, all PEM employees are required to display identification badges that must be worn at all times while in the facility. Visitors must sign in and out, are assigned a visitor badge, and are escorted by PEM personnel while at the facility. 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, PEM utilizes Random Access Control Facility 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, PEM 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.

ETS and PEM have implemented procedures and systems to provide efficient coordination of secure data exchange. This includes the established secure FTP site that is used for secure data transfers between ETS and PEM. These well-established procedures provide timely, efficient, and secure transfer of data. Access to the STAR data files is limited to appropriate personnel with direct project responsibilities.

Statistical Analysis

ETS Systems loads the PEM files in a database. The Data Quality Services 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 then keeps the files on secure servers and adheres to the ETS Code of Ethics to prevent any unauthorized access.

Reporting and Posting Results

After statistical analysis is completed for student results, the files flow into three different directions. Paper reports, some with individual student results and others with summary results, are produced. Encrypted files of summary results are also sent to CDE via FTP. Any summary results that have less than ten students are not reported. The statistics from the results are also entered into the item bank.

Student Confidentiality

To meet the requirements of the state and the *No Child Left Behind* Act of 2001, districts must collect demographic data about students, such as ethnicity, parent education, disabilities, whether the student qualifies for the school lunch program, etc. In addition, students may reveal other information about themselves through the essays they write. ETS takes precautions to prevent any of this information becoming public or being used other than for testing purposes. Such measures are applicable to all documents where these data may appear, including:

- Pre-ID files
- Reports
- Essays

Test Results

ETS also has security measures for files and reports that show students' scores and performance levels. ETS is committed to safeguarding the information in their possession from unauthorized access, disclosure, modification, or destruction. ETS has strict information security policies in order to protect the confidentiality of ETS and client data. ETS staff access to production databases is very limited. 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. Comprehensive disaster recovery facilities are

available and tested regularly at the SunGard installation in Philadelphia. ETS routinely sends backup data cartridges and files for critical software, applications, and documentation to an off-site storage facility for safekeeping to permit continued operation in the case of a disaster.

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 student's results on both electronic files and paper reports during:

- Scoring
- Transfer of scores via secure data exchange
- Reporting
- Erasure marks
- Internet postings
- 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 an ETS-contracted test, such as STAR. The ETS Office of Testing Integrity verifies that these standards are followed throughout the organization, including conducting periodic onsite security audits of departments, with follow-up reports containing recommendations for improvement.

References

Holland, P.W. and Thayer, D.T. "An alternative definition of the ETS delta scale of item difficulty." RR-85-43, 1985.

Dorans, N.J. and Holland, P.W. "DIF detection and description: Mantel-Haenszel and standardization," *Differential Item Functioning*. Edited by P. W. Holland and H. Wainer. Hillsdale, NJ: Erlbaum, pp. 35-66, 1993

Mantel, N. & Haenszel, W. "Statistical aspects of the analyses of data from retrospective studies of disease," *Journal of the National Cancer Institute*, Vol. 22, pp. 719-748, 1959.

Appendix 6.A—Demographic Summary Tables

Table 6.A.1 Demographic Summary for Reading/Language Arts Grade 2 (Overall Population)

	Number Tested	Mean Number Correct Score	Std. Dev. of Number Correct Score	Mean Percent Correct in Content Area				
				Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Overall Population Valid Scores	18,767	40	12	68%	60%	73%	64%	41%
Female	9,326	42	12	70%	64%	75%	66%	43%
Male	9,363	39	12	65%	57%	70%	61%	40%
Gender Unknown	78	39	12	66%	59%	70%	62%	39%
Argentina	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Bolivia	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brazil	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chile	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Colombia	11	43	7	69%	67%	77%	64%	48%
Costa Rica	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Cuba	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ecuador	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Spain	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Guatemala	89	37	13	62%	56%	65%	55%	39%
Mexico	2,226	38	13	63%	57%	68%	57%	39%
Panama	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Peru	21	45	12	76%	71%	77%	70%	42%
Puerto Rico	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Paraguay	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
El Salvador	166	36	13	59%	55%	66%	54%	37%
United States	2,835	41	12	68%	61%	74%	65%	42%
Uruguay	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Venezuela	5	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Other	151	40	13	66%	58%	72%	64%	42%
Country Unknown	13,233	41	12	68%	61%	73%	65%	42%
Not in NSLP	1,761	40	13	66%	60%	71%	62%	42%
In NSLP	16,700	41	12	68%	60%	73%	64%	42%
NSLP Unknown	306	38	12	63%	57%	67%	58%	38%
In U.S. Schools >= 12 Months	15,953	42	12	70%	62%	75%	67%	43%
In U.S. Schools < 12 Months	2,814	33	13	55%	49%	59%	48%	35%
EL in ELD	778	36	13	61%	55%	67%	56%	38%
EL in ELD and SDAIE	1,397	33	13	56%	50%	61%	50%	36%
EL in ELD and SDAIE with Primary Language Support	2,424	38	13	63%	56%	68%	58%	39%
EL in ELD and Academic Subjects through Primary Language	13,038	42	11	70%	62%	75%	67%	43%
Other EL Instructional Services	212	43	12	72%	65%	74%	68%	46%
None (EL only)	162	37	13	62%	55%	68%	58%	39%
Program Participation Unknown	756	41	12	69%	62%	73%	65%	42%
No Special Education	17,947	41	12	68%	61%	73%	64%	42%
Special Education	811	33	12	56%	48%	60%	52%	35%
Special Education Unknown	9	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 6.A.2 Demographic Summary for Reading/Language Arts Grade 2 (Target Population)

	Number Tested	Mean Number Correct Score	Std. Dev. of Number Correct Score	Mean Percent Correct in Content Area				
				Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target Population Valid Scores	12,486	41	12	68%	61%	73%	65%	42%
Female	6,163	42	12	70%	64%	76%	67%	43%
Male	6,323	39	12	66%	57%	71%	62%	40%
Gender Unknown	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Argentina	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Bolivia	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brazil	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chile	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Colombia	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Costa Rica	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Cuba	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ecuador	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Spain	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Guatemala	52	40	12	65%	60%	70%	62%	41%
Mexico	612	41	13	67%	61%	72%	64%	42%
Panama	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Peru	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Puerto Rico	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Paraguay	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
El Salvador	84	41	12	67%	63%	72%	65%	41%
United States	983	41	12	68%	60%	73%	65%	41%
Uruguay	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Venezuela	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Other	111	43	11	71%	63%	77%	70%	44%
Country Unknown	10,620	41	12	68%	61%	73%	65%	42%
Not in NSLP	1,133	40	13	67%	60%	71%	64%	42%
In NSLP	11,324	41	12	68%	61%	73%	65%	42%
NSLP Unknown	29	38	9	66%	59%	70%	54%	38%
In U.S. Schools >= 12 Months	10,588	42	11	70%	62%	75%	67%	43%
In U.S. Schools < 12 Months	1,898	33	13	56%	50%	59%	50%	35%
EL in ELD	165	32	12	53%	48%	57%	47%	33%
EL in ELD and SDAIE	576	31	12	52%	47%	56%	45%	34%
EL in ELD and SDAIE with Primary Language Support	629	30	13	51%	46%	54%	44%	32%
EL in ELD and Academic Subjects through Primary Language	11,008	42	11	70%	62%	75%	67%	43%
Other EL Instructional Services	12	37	16	60%	58%	61%	57%	41%
None (EL only)	73	33	13	55%	48%	58%	50%	36%
Program Participation Unknown	23	34	12	57%	57%	53%	49%	39%
No Special Education	11,879	41	12	69%	61%	74%	65%	42%
Special Education	607	34	12	58%	49%	62%	54%	36%
Special Education Unknown	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 6.A.3 Demographic Summary for Reading/Language Arts Grade 3 (Overall Population)

	Number Tested	Mean Number Correct Score	Std. Dev. of Number Correct Score	Mean Percent Correct in Content Area				
				Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Overall Population Valid Scores	13,084	37	12	67%	50%	54%	53%	49%
Female	6,512	38	11	70%	53%	57%	56%	52%
Male	6,538	35	11	65%	47%	52%	50%	46%
Gender Unknown	34	31	10	60%	44%	47%	38%	41%
Argentina	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Bolivia	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brazil	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chile	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Colombia	7	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Costa Rica	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Cuba	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ecuador	10	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Spain	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Guatemala	109	34	13	64%	48%	49%	49%	45%
Mexico	1,716	34	12	64%	47%	52%	48%	44%
Panama	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Peru	15	44	12	79%	59%	68%	64%	59%
Puerto Rico	9	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Paraguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
El Salvador	163	34	11	64%	47%	55%	47%	42%
United States	1,854	37	11	67%	51%	55%	55%	50%
Uruguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Venezuela	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Other	114	36	11	68%	50%	53%	52%	48%
Country Unknown	9,074	37	11	68%	51%	54%	54%	50%
Not in NSLP	1,364	37	12	68%	51%	56%	54%	50%
In NSLP	11,543	37	11	67%	50%	54%	53%	49%
NSLP Unknown	177	32	11	60%	44%	49%	45%	41%
In U.S. Schools >= 12 Months	10,482	38	11	69%	51%	55%	55%	51%
In U.S. Schools < 12 Months	2,602	33	11	62%	45%	51%	45%	41%
EL in ELD	600	33	11	61%	44%	50%	46%	41%
EL in ELD and SDAIE	1,324	34	12	63%	45%	52%	47%	43%
EL in ELD and SDAIE with Primary Language Support	1,714	35	12	64%	48%	53%	50%	46%
EL in ELD and Academic Subjects through Primary Language	8,637	38	11	69%	52%	55%	56%	51%
Other EL Instructional Services	194	40	12	72%	56%	60%	58%	55%
None (EL only)	121	35	12	65%	48%	54%	50%	47%
Program Participation Unknown	494	36	13	65%	49%	53%	52%	48%
No Special Education	12,459	37	11	68%	51%	55%	54%	50%
Special Education	611	30	11	55%	40%	45%	43%	38%
Special Education Unknown	14	30	11	56%	40%	45%	44%	34%

Table 6.A.4 Demographic Summary for Reading/Language Arts Grade 3 (Target Population)

	Number Tested	Mean Number Correct Score	Std. Dev. of Number Correct Score	Mean Percent Correct in Content Area				
				Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target Population Valid Scores	8,952	37	11	68%	51%	55%	54%	50%
Female	4,471	39	11	70%	53%	57%	57%	52%
Male	4,481	35	11	66%	48%	53%	51%	47%
Gender Unknown	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Argentina	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Bolivia	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brazil	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chile	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Colombia	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Costa Rica	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Cuba	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ecuador	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Spain	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Guatemala	62	37	12	68%	52%	53%	53%	51%
Mexico	549	37	11	69%	51%	57%	53%	50%
Panama	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Peru	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Puerto Rico	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Paraguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
El Salvador	78	35	11	66%	48%	57%	49%	44%
United States	805	37	11	68%	50%	55%	54%	49%
Uruguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Venezuela	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Other	83	36	11	68%	50%	51%	52%	50%
Country Unknown	7,360	37	11	68%	51%	54%	54%	50%
Not in NSLP	942	38	12	69%	52%	57%	55%	50%
In NSLP	7,989	37	11	68%	50%	54%	54%	50%
NSLP Unknown	21	37	9	67%	52%	51%	54%	56%
In U.S. Schools >= 12 Months	7,243	38	11	69%	52%	55%	56%	52%
In U.S. Schools < 12 Months	1,709	33	11	63%	45%	51%	46%	42%
EL in ELD	154	32	12	59%	42%	50%	44%	39%
EL in ELD and SDAIE	582	33	12	62%	45%	51%	46%	41%
EL in ELD and SDAIE with Primary Language Support	577	31	11	59%	42%	49%	43%	39%
EL in ELD and Academic Subjects through Primary Language	7,555	38	11	69%	52%	55%	56%	51%
Other EL Instructional Services	11	33	15	61%	49%	50%	40%	48%
None (EL only)	63	33	10	64%	46%	53%	44%	41%
Program Participation Unknown	10	N/A	N/A	N/A	N/A	N/A	N/A	N/A
No Special Education	8,484	37	11	69%	51%	55%	55%	50%
Special Education	468	30	11	55%	40%	46%	44%	39%
Special Education Unknown	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 6.A.5 Demographic Summary for Reading/Language Arts Grade 4 (Overall Population)

	Number Tested	Mean Number Correct Score	Std. Dev. of Number Correct Score	Mean Percent Correct in Content Area				
				Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Overall Population Valid Scores	7,972	40	14	59%	50%	43%	58%	51%
Female	4,046	43	13	63%	53%	46%	62%	54%
Male	3,896	38	14	56%	47%	40%	54%	48%
Gender Unknown	30	38	13	57%	45%	35%	55%	51%
Argentina	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Bolivia	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brazil	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chile	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Colombia	10	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Costa Rica	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Cuba	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ecuador	5	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Spain	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Guatemala	62	40	14	58%	50%	45%	56%	50%
Mexico	1,309	38	13	56%	48%	41%	54%	49%
Panama	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Peru	11	46	13	68%	59%	43%	70%	58%
Puerto Rico	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Paraguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
El Salvador	118	40	13	59%	51%	42%	57%	51%
United States	973	40	14	59%	49%	42%	59%	51%
Uruguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Venezuela	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Other	91	39	14	60%	48%	42%	55%	49%
Country Unknown	5,370	41	14	60%	51%	44%	59%	52%
Not in NSLP	892	41	14	60%	52%	45%	60%	53%
In NSLP	6,939	40	14	59%	50%	43%	58%	51%
NSLP Unknown	141	36	12	55%	44%	37%	52%	46%
In U.S. Schools >= 12 Months	5,532	41	14	60%	50%	43%	59%	52%
In U.S. Schools < 12 Months	2,440	39	13	57%	49%	43%	55%	50%
EL in ELD	511	38	13	56%	48%	42%	53%	48%
EL in ELD and SDAIE	1,214	39	13	57%	49%	43%	54%	49%
EL in ELD and SDAIE with Primary Language Support	1,419	39	13	57%	49%	43%	56%	50%
EL in ELD and Academic Subjects through Primary Language	4,413	41	14	61%	51%	44%	60%	52%
Other EL Instructional Services	91	42	16	63%	53%	48%	59%	54%
None (EL only)	62	43	13	63%	56%	45%	60%	56%
Program Participation Unknown	262	42	14	61%	55%	46%	60%	54%
No Special Education	7,655	41	13	60%	51%	44%	59%	52%
Special Education	305	29	12	42%	36%	33%	42%	37%
Special Education Unknown	12	32	11	50%	38%	29%	47%	39%

Table 6.A.6 Demographic Summary for Reading/Language Arts Grade 4 (Target Population)

	Number Tested	Mean Number Correct Score	Std. Dev. of Number Correct Score	Mean Percent Correct in Content Area				
				Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target Population Valid Scores	5,291	40	14	60%	50%	43%	59%	52%
Female	2,676	43	13	63%	53%	47%	63%	55%
Male	2,615	38	13	56%	47%	40%	54%	48%
Gender Unknown	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Argentina	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Bolivia	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brazil	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chile	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Colombia	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Costa Rica	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Cuba	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ecuador	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Spain	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Guatemala	23	36	14	55%	44%	42%	49%	49%
Mexico	351	39	13	59%	48%	42%	56%	49%
Panama	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Peru	5	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Puerto Rico	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Paraguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
El Salvador	46	40	12	60%	50%	43%	59%	52%
United States	481	39	13	59%	47%	40%	58%	50%
Uruguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Venezuela	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Other	47	41	14	62%	50%	44%	59%	53%
Country Unknown	4,329	41	14	60%	50%	44%	59%	52%
Not in NSLP	565	41	14	61%	52%	45%	60%	53%
In NSLP	4,704	40	14	60%	50%	43%	59%	52%
NSLP Unknown	22	33	13	52%	40%	24%	50%	41%
In U.S. Schools >= 12 Months	3,718	41	14	61%	50%	43%	60%	52%
In U.S. Schools < 12 Months	1,573	39	13	58%	49%	44%	56%	50%
EL in ELD	163	38	13	57%	48%	44%	54%	50%
EL in ELD and SDAIE	590	39	14	58%	49%	44%	54%	49%
EL in ELD and SDAIE with Primary Language Support	578	39	13	58%	49%	43%	56%	49%
EL in ELD and Academic Subjects through Primary Language	3,896	41	14	61%	50%	43%	60%	52%
Other EL Instructional Services	5	N/A	N/A	N/A	N/A	N/A	N/A	N/A
None (EL only)	43	41	12	60%	53%	44%	58%	53%
Program Participation Unknown	16	36	12	53%	46%	36%	51%	44%
No Special Education	5,073	41	13	61%	51%	44%	59%	52%
Special Education	218	29	13	43%	37%	32%	42%	38%
Special Education Unknown	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 6.A.7 Demographic Summary for Mathematics Grade 2 (Overall Population)

	Number Tested	Mean Number Correct Score	Std. Dev. of Number Correct Score	Mean Percent Correct in Content Area				
				Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Overall Population Valid Scores	18,665	44	12	67%	66%	63%	75%	71%
Female	9,269	45	11	67%	67%	62%	74%	71%
Male	9,317	44	12	66%	65%	64%	75%	71%
Gender Unknown	79	43	12	64%	65%	64%	73%	63%
Argentina	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Bolivia	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brazil	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chile	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Colombia	11	48	13	72%	71%	71%	78%	81%
Costa Rica	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Cuba	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ecuador	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Spain	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Guatemala	89	41	13	61%	64%	58%	69%	63%
Mexico	2,224	42	12	62%	63%	60%	71%	66%
Panama	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Peru	21	48	11	80%	68%	76%	76%	78%
Puerto Rico	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Paraguay	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
El Salvador	166	41	12	59%	63%	56%	69%	65%
United States	2,818	45	11	68%	67%	64%	75%	72%
Uruguay	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Venezuela	5	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Other	151	44	12	66%	66%	60%	75%	70%
Country Unknown	13,151	45	11	67%	66%	64%	75%	72%
Not in NSLP	1,765	44	12	66%	66%	63%	75%	70%
In NSLP	16,594	45	11	67%	66%	63%	75%	71%
NSLP Unknown	306	42	12	61%	62%	60%	71%	65%
In U.S. Schools >= 12 Months	15,837	46	11	69%	68%	65%	76%	73%
In U.S. Schools < 12 Months	2,828	38	12	55%	58%	55%	67%	58%
EL in ELD	779	40	12	60%	60%	58%	69%	62%
EL in ELD and SDAIE	1,409	39	12	57%	59%	56%	68%	61%
EL in ELD and SDAIE with Primary Language Support	2,363	42	12	63%	63%	61%	72%	67%
EL in ELD and Academic Subjects through Primary Language	12,994	46	11	69%	68%	65%	76%	73%
Other EL Instructional Services	209	45	12	67%	66%	64%	76%	72%
None (EL only)	161	40	12	60%	59%	58%	70%	63%
Program Participation Unknown	750	45	12	68%	67%	65%	75%	71%
No Special Education	17,838	45	11	67%	67%	64%	75%	71%
Special Education	818	38	12	58%	56%	54%	66%	61%
Special Education Unknown	9	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 6.A.8 Demographic Summary for Mathematics Grade 2 (Target Population)

	Number Tested	Mean Number Correct Score	Std. Dev. of Number Correct Score	Mean Percent Correct in Content Area				
				Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target Population Valid Scores	12,469	45	11	68%	67%	64%	75%	72%
Female	6,153	45	11	68%	67%	62%	74%	72%
Male	6,316	45	11	67%	66%	65%	76%	72%
Gender Unknown	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Argentina	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Bolivia	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brazil	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chile	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Colombia	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Costa Rica	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Cuba	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ecuador	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Spain	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Guatemala	52	45	13	68%	70%	62%	71%	68%
Mexico	615	45	12	68%	68%	65%	75%	72%
Panama	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Peru	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Puerto Rico	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Paraguay	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
El Salvador	85	45	11	67%	71%	65%	74%	72%
United States	985	45	11	69%	67%	64%	74%	72%
Uruguay	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Venezuela	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Other	111	47	11	70%	71%	64%	77%	75%
Country Unknown	10,597	45	11	67%	67%	64%	75%	72%
Not in NSLP	1,133	45	12	67%	67%	63%	76%	70%
In NSLP	11,307	45	11	68%	67%	64%	75%	72%
NSLP Unknown	29	45	8	62%	70%	74%	77%	72%
In U.S. Schools >= 12 Months	10,563	46	11	69%	68%	65%	76%	74%
In U.S. Schools < 12 Months	1,906	39	13	57%	59%	56%	68%	60%
EL in ELD	165	37	12	53%	55%	54%	66%	53%
EL in ELD and SDAIE	583	38	12	54%	57%	54%	67%	58%
EL in ELD and SDAIE with Primary Language Support	632	38	12	54%	57%	53%	66%	56%
EL in ELD and Academic Subjects through Primary Language	10,982	46	11	69%	68%	65%	76%	74%
Other EL Instructional Services	13	33	14	48%	49%	53%	57%	52%
None (EL only)	72	39	12	58%	58%	56%	69%	59%
Program Participation Unknown	22	42	13	64%	62%	52%	74%	62%
No Special Education	11,854	45	11	68%	67%	64%	76%	72%
Special Education	615	39	12	59%	57%	56%	67%	63%
Special Education Unknown	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 6.A.9 Demographic Summary for Mathematics Grade 3 (Overall Population)

	Number Tested	Mean Number Correct Score	Std. Dev. of Number Correct Score	Mean Percent Correct in Content Area				
				Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Overall Population Valid Scores	13,031	43	13	64%	67%	61%	69%	74%
Female	6,479	43	12	64%	68%	60%	70%	76%
Male	6,519	43	13	64%	66%	61%	69%	73%
Gender Unknown	33	40	14	58%	64%	54%	65%	67%
Argentina	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Bolivia	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brazil	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chile	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Colombia	7	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Costa Rica	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Cuba	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ecuador	10	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Spain	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Guatemala	110	39	15	58%	63%	55%	62%	66%
Mexico	1,704	39	13	58%	62%	55%	64%	66%
Panama	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Peru	15	50	11	72%	77%	72%	81%	89%
Puerto Rico	9	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Paraguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
El Salvador	162	37	13	53%	57%	50%	62%	63%
United States	1,836	45	12	67%	69%	64%	72%	78%
Uruguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Venezuela	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Other	114	42	12	62%	66%	60%	69%	72%
Country Unknown	9,050	44	13	65%	68%	61%	70%	76%
Not in NSLP	1,359	43	13	65%	68%	60%	69%	74%
In NSLP	11,497	43	13	64%	67%	61%	70%	75%
NSLP Unknown	175	38	14	56%	60%	52%	61%	61%
In U.S. Schools >= 12 Months	10,430	45	12	67%	69%	63%	72%	78%
In U.S. Schools < 12 Months	2,601	37	13	55%	60%	51%	60%	60%
EL in ELD	599	38	13	56%	60%	52%	61%	62%
EL in ELD and SDAIE	1,331	39	13	59%	63%	54%	62%	65%
EL in ELD and SDAIE with Primary Language Support	1,694	40	14	59%	64%	55%	65%	66%
EL in ELD and Academic Subjects through Primary Language	8,605	45	12	67%	69%	63%	72%	78%
Other EL Instructional Services	188	46	13	70%	71%	64%	73%	80%
None (EL only)	122	41	14	62%	64%	56%	65%	69%
Program Participation Unknown	492	42	14	63%	65%	59%	68%	71%
No Special Education	12,407	43	13	65%	68%	61%	70%	75%
Special Education	610	37	13	56%	57%	52%	61%	66%
Special Education Unknown	14	35	13	51%	58%	52%	56%	59%

Table 6.A.10 Demographic Summary for Mathematics Grade 3 (Target Population)

	Number Tested	Mean Number Correct Score	Std. Dev. of Number Correct Score	Mean Percent Correct in Content Area				
				Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target Population Valid Scores	8,926	44	12	65%	68%	62%	70%	76%
Female	4,454	44	12	65%	69%	61%	71%	78%
Male	4,472	44	13	65%	67%	62%	70%	74%
Gender Unknown	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Argentina	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Bolivia	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brazil	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chile	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Colombia	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Costa Rica	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Cuba	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ecuador	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Spain	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Guatemala	62	44	15	64%	69%	63%	70%	79%
Mexico	540	43	13	63%	68%	60%	69%	74%
Panama	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Peru	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Puerto Rico	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Paraguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
El Salvador	78	40	13	59%	61%	55%	68%	67%
United States	797	45	12	67%	69%	65%	72%	78%
Uruguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Venezuela	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Other	83	42	12	62%	64%	59%	71%	74%
Country Unknown	7,351	44	12	65%	68%	62%	70%	76%
Not in NSLP	941	44	13	66%	69%	62%	70%	76%
In NSLP	7,965	44	12	65%	68%	62%	70%	76%
NSLP Unknown	20	47	10	69%	65%	68%	78%	87%
In U.S. Schools >= 12 Months	7,215	45	12	67%	70%	64%	73%	79%
In U.S. Schools < 12 Months	1,711	38	13	56%	61%	52%	61%	61%
EL in ELD	154	36	13	53%	60%	50%	59%	58%
EL in ELD and SDAIE	587	38	13	57%	62%	53%	61%	63%
EL in ELD and SDAIE with Primary Language Support	573	36	13	53%	59%	50%	59%	57%
EL in ELD and Academic Subjects through Primary Language	7,526	45	12	67%	69%	64%	73%	79%
Other EL Instructional Services	12	36	14	56%	61%	43%	54%	63%
None (EL only)	64	37	14	55%	59%	51%	60%	60%
Program Participation Unknown	10	N/A	N/A	N/A	N/A	N/A	N/A	N/A
No Special Education	8,464	44	12	66%	69%	62%	71%	76%
Special Education	462	37	13	56%	57%	53%	61%	67%
Special Education Unknown	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 6.A.11 Demographic Summary for Mathematics Grade 4 (Overall Population)

	Number Tested	Mean Number Correct Score	Std. Dev. of Number Correct Score	Mean Percent Correct in Content Area				
				Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Overall Population Valid Scores	7,943	39	13	64%	56%	62%	60%	53%
Female	4,038	40	13	64%	57%	64%	62%	54%
Male	3,875	38	14	64%	55%	60%	58%	52%
Gender Unknown	30	36	13	56%	54%	57%	56%	55%
Argentina	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Bolivia	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brazil	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chile	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Colombia	10	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Costa Rica	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Cuba	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ecuador	5	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Spain	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Guatemala	62	35	14	54%	50%	56%	59%	46%
Mexico	1,299	35	13	57%	51%	54%	55%	46%
Panama	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Peru	10	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Puerto Rico	4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Paraguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
El Salvador	117	36	13	58%	49%	56%	59%	47%
United States	961	41	13	66%	58%	65%	63%	56%
Uruguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Venezuela	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Other	91	38	13	63%	52%	58%	59%	54%
Country Unknown	5,365	40	13	66%	57%	64%	61%	54%
Not in NSLP	891	40	13	65%	57%	62%	61%	54%
In NSLP	6,912	39	13	64%	56%	62%	60%	53%
NSLP Unknown	140	34	12	55%	47%	53%	55%	44%
In U.S. Schools >= 12 Months	5,506	41	13	67%	58%	65%	62%	56%
In U.S. Schools < 12 Months	2,437	35	13	58%	52%	54%	55%	47%
EL in ELD	509	35	13	59%	50%	55%	55%	46%
EL in ELD and SDAIE	1,212	36	13	59%	53%	55%	55%	48%
EL in ELD and SDAIE with Primary Language Support	1,414	37	13	60%	53%	58%	57%	49%
EL in ELD and Academic Subjects through Primary Language	4,402	41	13	68%	58%	66%	63%	56%
Other EL Instructional Services	91	41	15	64%	58%	66%	62%	57%
None (EL only)	62	38	13	61%	55%	59%	57%	48%
Program Participation Unknown	253	40	14	66%	57%	64%	60%	58%
No Special Education	7,621	40	13	65%	57%	62%	61%	53%
Special Education	310	32	13	53%	46%	49%	50%	47%
Special Education Unknown	12	31	11	51%	40%	47%	51%	44%

Table 6.A.12 Demographic Summary for Mathematics Grade 4 (Target Population)

	Number Tested	Mean Number Correct Score	Std. Dev. of Number Correct Score	Mean Percent Correct in Content Area				
				Word Analysis and Vocabulary Development	Reading Comprehension	Literary Response and Analysis	Written Conventions	Writing Strategies
Target Population Valid Scores	5,283	40	13	65%	57%	63%	61%	54%
Female	2,676	41	13	65%	58%	65%	63%	56%
Male	2,607	39	13	66%	55%	61%	59%	53%
Gender Unknown	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Argentina	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Bolivia	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brazil	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chile	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Colombia	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Costa Rica	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Cuba	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ecuador	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Spain	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Guatemala	23	36	15	58%	49%	57%	62%	47%
Mexico	352	37	13	61%	53%	58%	61%	48%
Panama	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Peru	5	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Puerto Rico	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Paraguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
El Salvador	45	38	12	61%	49%	60%	63%	49%
United States	481	41	12	66%	56%	65%	64%	57%
Uruguay	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Venezuela	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Other	46	43	13	72%	58%	67%	67%	59%
Country Unknown	4,322	40	13	66%	57%	63%	61%	54%
Not in NSLP	565	40	13	66%	58%	62%	62%	54%
In NSLP	4,696	40	13	65%	57%	63%	61%	54%
NSLP Unknown	22	34	12	59%	44%	51%	56%	49%
In U.S. Schools >= 12 Months	3,711	42	13	68%	58%	66%	63%	57%
In U.S. Schools < 12 Months	1,572	36	13	59%	53%	56%	56%	48%
EL in ELD	163	36	13	59%	52%	56%	55%	46%
EL in ELD and SDAIE	591	35	13	59%	53%	54%	54%	48%
EL in ELD and SDAIE with Primary Language Support	577	36	12	58%	51%	55%	56%	47%
EL in ELD and Academic Subjects through Primary Language	3,888	42	13	68%	58%	66%	63%	57%
Other EL Instructional Services	5	N/A	N/A	N/A	N/A	N/A	N/A	N/A
None (EL only)	43	36	12	59%	54%	59%	56%	41%
Program Participation Unknown	16	34	13	59%	54%	45%	51%	47%
No Special Education	5,061	40	13	66%	57%	64%	62%	54%
Special Education	222	33	13	55%	47%	50%	51%	48%
Special Education Unknown	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Appendix 6.B—Test Variations, Accommodations, and Modifications

Test Variation (1) / Accommodation (2) / Modification (3) *	Provision
A. Test administration directions that are simplified or clarified (does not apply to test questions)	All
Test individual student separately, provided that a test examiner directly supervises the student	1
Visual magnifying equipment	1
Audio amplification equipment	1
Noise buffers (e.g. individual carrel or study enclosure)	1
Special lighting or acoustics; special or adaptive furniture	1
Colored overlay, mask, or other means to maintain visual attention	1
Student marks in test booklet (other than responses)	All (For grades 2, 3 – marks must be erased to avoid scanning interference)
B. 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	2
C. Responses dictated [orally, or in Manually Coded English or American Sign Language] to a scribe for selected-response items (multiple-choice questions)	2
F. 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)	2
G. Braille transcriptions provided by the test contractor	2
H. Large-print versions [Test items enlarged if font larger than required on large-print versions]	2
I. Extra time on a test within a testing day	All
J. Test over more than one day for a test or test part to be administered in a single sitting	2
K. Supervised breaks within a section of the test	2
L. Administration of the test at the most beneficial time of day to the student	2
M. Test administered at home or in hospital by a test examiner	2
N. Dictionary	3
P. Test questions read aloud to student or audio presentation (CD)	2 (Math) 3 (Reading, Language, Spelling)
Q. Calculators on the mathematics or science tests	3
R. Arithmetic table on the mathematics or science tests	3
S. Math manipulatives on the mathematics or science tests	3
V. Assistive device that interferes with the independent work of the student on the multiple-choice and/or essay responses	3

* All = These test variations may be provided to all students.

Test Variation (1) = Eligible students may have testing variations if regularly used in the classroom.

Accommodation (2) = Eligible students shall be permitted to take the examination/test with accommodations if specified in the eligible student's IEP or Section 504 Plan for use on the examination, standardized testing, or for use during classroom instruction and assessment.

Modification (3) = Eligible students shall be permitted to take the tests with modifications if specified in the eligible student's IEP or Section 504 Plan.

Appendix 6.C — Accommodation Summary Tables

Table 6.C.1 Accommodation Summary for Reading/Language Arts Grades 2, 3, and 4

Accommodation Summary for Reading/Language Arts Grades 2, 3, and 4 ³						
All Students Tested	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	1	0.01%	0	0.00%	3	0.04%
C: Dictated responses to a scribe	2	0.01%	1	0.01%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	1	0.01%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	3	0.02%	0	0.00%
J: Tested over more than one day	17	0.09%	4	0.03%	3	0.04%
K: Had supervised breaks	37	0.20%	40	0.30%	19	0.24%
L: Most beneficial time of day	18	0.09%	21	0.16%	16	0.20%
M: Administered at home or in a hospital	1	0.01%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	13	0.07%	13	0.10%	4	0.05%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	3	0.02%	0	0.00%
X: Used an unlisted accommodation	12	0.06%	4	0.03%	8	0.10%
Accom. or Mod. is in Section 504 Plan	3	0.02%	2	0.02%	1	0.01%
Accom. or Mod. is in IEP	53	0.28%	54	0.41%	35	0.44%
Any Accom., Mod., or EL Variation ⁴	63	0.33%	62	0.47%	40	0.50%
No Accom., Mod., or EL Variation	18,885	99.67%	13,163	99.53%	7,965	99.50%

Target Students Tested	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	1	0.01%	0	0.00%	0	0.00%
C: Dictated responses to a scribe	2	0.02%	1	0.01%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	1	0.01%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	11	0.09%	3	0.03%	2	0.04%
K: Had supervised breaks	26	0.21%	23	0.25%	8	0.15%
L: Most beneficial time of day	9	0.07%	14	0.16%	11	0.21%
M: Administered at home or in a hospital	1	0.01%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	13	0.10%	4	0.04%	3	0.06%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	2	0.02%	0	0.00%
X: Used an unlisted accommodation	7	0.06%	1	0.01%	5	0.09%
Accom. or Mod. is in Section 504 Plan	2	0.02%	1	0.01%	1	0.02%
Accom. or Mod. is in IEP	39	0.31%	30	0.33%	20	0.38%
Any Accom., Mod., or EL Variation	45	0.36%	33	0.37%	22	0.41%
No Accom., Mod., or EL Variation	12,546	99.64%	8,996	99.63%	5,290	99.59%

³ The sum of the numbers of students across subgroups may not match exactly to the total testing population. For example, Students in Special Education + Students not in Special Education \neq All Tested for the provision of an accommodation. This occurred due to the fact that only valid Primary Disability codes were chosen to identify those subgroups.

⁴ There are some students who have multiple accommodations and modifications, so the number of students in “Any Accom., Mod., or EL Variation” does not equal the sum of the students from each accommodation or modification category.

Accommodation Summary for Reading/Language Arts Grades 2, 3, and 4³

Non-Target (Optional) Students Tested	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	0	0.00%	0	0.00%	3	0.11%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	3	0.07%	0	0.00%
J: Tested over more than one day	6	0.09%	1	0.02%	1	0.04%
K: Had supervised breaks	11	0.17%	17	0.41%	11	0.41%
L: Most beneficial time of day	9	0.14%	7	0.17%	5	0.19%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	0	0.00%	9	0.21%	1	0.04%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	1	0.02%	0	0.00%
X: Used an unlisted accommodation	5	0.08%	3	0.07%	3	0.11%
Accom. or Mod. is in Section 504 Plan	1	0.02%	1	0.02%	0	0.00%
Accom. or Mod. is in IEP	14	0.22%	24	0.57%	15	0.56%
Any Accom., Mod., or EL Variation	18	0.28%	29	0.69%	18	0.67%
No Accom., Mod., or EL Variation	6,339	99.72%	4,167	99.31%	2,675	99.33%

Students Not in Special Education	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	0	0.00%	0	0.00%	3	0.04%
C: Dictated responses to a scribe	2	0.01%	1	0.01%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	1	0.01%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	1	0.01%	0	0.00%
J: Tested over more than one day	3	0.02%	1	0.01%	0	0.00%
K: Had supervised breaks	8	0.04%	10	0.08%	5	0.07%
L: Most beneficial time of day	4	0.02%	11	0.09%	5	0.07%
M: Administered at home or in a hospital	1	0.01%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	3	0.02%	3	0.02%	1	0.01%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	3	0.02%	2	0.02%	0	0.00%
Accom. or Mod. is in Section 504 Plan	3	0.02%	2	0.02%	1	0.01%
Accom. or Mod. is in IEP	10	0.06%	11	0.09%	5	0.07%
Any Accom., Mod., or EL Variation	15	0.08%	17	0.14%	10	0.13%
No Accom., Mod., or EL Variation	18,099	99.92%	12,566	99.86%	7,672	99.87%

Students in Special Education	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	1	0.12%	0	0.00%	0	0.00%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	1	0.16%	0	0.00%
J: Tested over more than one day	14	1.70%	3	0.48%	3	0.96%
K: Had supervised breaks	29	3.52%	30	4.78%	10	3.22%
L: Most beneficial time of day	14	1.70%	10	1.59%	6	1.93%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%

Accommodation Summary for Reading/Language Arts Grades 2, 3, and 4³

N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	10	1.21%	10	1.59%	3	0.96%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	2	0.32%	0	0.00%
X: Used an unlisted accommodation	9	1.09%	2	0.32%	8	2.57%
Accom. or Mod. is in Section 504 Plan	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in IEP	43	5.21%	41	6.54%	25	8.04%
Any Accom., Mod., or EL Variation	48	5.82%	43	6.86%	25	8.04%
No Accom., Mod., or EL Variation	777	94.18%	584	93.14%	286	91.96%

Students in U.S. Schools < 12 Months	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	0	0.00%	0	0.00%	0	0.00%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	1	0.03%	0	0.00%	0	0.00%
K: Had supervised breaks	1	0.03%	5	0.19%	1	0.04%
L: Most beneficial time of day	0	0.00%	4	0.15%	2	0.08%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	0	0.00%	3	0.11%	1	0.04%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	0	0.00%	1	0.04%	0	0.00%
Accom. or Mod. is in Section 504 Plan	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in IEP	1	0.03%	4	0.15%	2	0.08%
Any Accom., Mod., or EL Variation	1	0.03%	7	0.26%	3	0.12%
No Accom., Mod., or EL Variation	2,863	99.97%	2,652	99.74%	2,450	99.88%

Students in U.S. Schools >= 12 Months	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	1	0.01%	0	0.00%	3	0.05%
C: Dictated responses to a scribe	2	0.01%	1	0.01%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	1	0.01%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	3	0.03%	0	0.00%
J: Tested over more than one day	16	0.10%	4	0.04%	3	0.05%
K: Had supervised breaks	36	0.22%	35	0.33%	18	0.32%
L: Most beneficial time of day	18	0.11%	17	0.16%	14	0.25%
M: Administered at home or in a hospital	1	0.01%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	13	0.08%	10	0.09%	3	0.05%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	3	0.03%	0	0.00%
X: Used an unlisted accommodation	12	0.07%	3	0.03%	8	0.14%
Accom. or Mod. is in Section 504 Plan	3	0.02%	2	0.02%	1	0.02%
Accom. or Mod. is in IEP	52	0.32%	50	0.47%	33	0.59%
Any Accom., Mod., or EL Variation	62	0.39%	55	0.52%	37	0.67%
No Accom., Mod., or EL Variation	16,022	99.61%	10,511	99.48%	5,515	99.33%

Accommodation Summary for Reading/Language Arts Grades 2, 3, and 4³

EL Program: EL in ELD	Grade 2		Grade 3		Grade 4	
	Count	Pct. of	Count	Pct. of	Count	Pct. of
		Total		Total		Total
B: Marked in test booklet	0	0.00%	0	0.00%	0	0.00%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	0	0.00%	0	0.00%	0	0.00%
K: Had supervised breaks	0	0.00%	0	0.00%	0	0.00%
L: Most beneficial time of day	0	0.00%	0	0.00%	1	0.19%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	0	0.00%	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	3	0.38%	3	0.49%	3	0.58%
Accom. or Mod. is in Section 504 Plan	0	0.00%	1	0.16%	0	0.00%
Accom. or Mod. is in IEP	3	0.38%	2	0.33%	3	0.58%
Any Accom., Mod., or EL Variation	3	0.38%	3	0.49%	4	0.78%
No Accom., Mod., or EL Variation	789	99.62%	604	99.51%	510	99.22%

EL Program: EL in ELD and SDAIE	Grade 2		Grade 3		Grade 4	
	Count	Pct. of	Count	Pct. of	Count	Pct. of
		Total		Total		Total
B: Marked in test booklet	0	0.00%	0	0.00%	0	0.00%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	3	0.21%	1	0.07%	1	0.08%
K: Had supervised breaks	3	0.21%	7	0.51%	4	0.33%
L: Most beneficial time of day	3	0.21%	2	0.15%	0	0.00%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	0	0.00%	1	0.07%	2	0.16%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	0	0.00%	1	0.07%	0	0.00%
Accom. or Mod. is in Section 504 Plan	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in IEP	3	0.21%	6	0.44%	5	0.41%
Any Accom., Mod., or EL Variation	3	0.21%	8	0.59%	5	0.41%
No Accom., Mod., or EL Variation	1,415	99.79%	1,355	99.41%	1,221	99.59%

EL Program: EL in ELD and SDAIE with Primary Language Support	Grade 2		Grade 3		Grade 4	
	Count	Pct. of	Count	Pct. of	Count	Pct. of
		Total		Total		Total
B: Marked in test booklet	0	0.00%	0	0.00%	0	0.00%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	1	0.04%	0	0.00%	0	0.00%
K: Had supervised breaks	1	0.04%	2	0.11%	3	0.21%
L: Most beneficial time of day	1	0.04%	1	0.06%	0	0.00%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%

Accommodation Summary for Reading/Language Arts Grades 2, 3, and 4³

N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	0	0.00%	1	0.06%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in Section 504 Plan	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in IEP	1	0.04%	2	0.11%	3	0.21%
Any Accom., Mod., or EL Variation	1	0.04%	2	0.11%	3	0.21%
No Accom., Mod., or EL Variation	2,457	99.96%	1,738	99.89%	1,418	99.79%

EL Program: EL in ELD and Academic Subjects through Primary Language	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	1	0.01%	0	0.00%	2	0.05%
C: Dictated responses to a scribe	2	0.02%	1	0.01%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	1	0.01%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	2	0.02%	0	0.00%
J: Tested over more than one day	13	0.10%	3	0.03%	2	0.05%
K: Had supervised breaks	32	0.24%	29	0.33%	8	0.18%
L: Most beneficial time of day	13	0.10%	17	0.20%	10	0.23%
M: Administered at home or in a hospital	1	0.01%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	13	0.10%	9	0.10%	2	0.05%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	2	0.02%	0	0.00%
X: Used an unlisted accommodation	9	0.07%	0	0.00%	5	0.11%
Accom. or Mod. is in Section 504 Plan	3	0.02%	1	0.01%	1	0.02%
Accom. or Mod. is in IEP	45	0.34%	40	0.46%	19	0.43%
Any Accom., Mod., or EL Variation	55	0.42%	44	0.51%	22	0.50%
No Accom., Mod., or EL Variation	13,077	99.58%	8,652	99.49%	4,405	99.50%

EL Program: Other EL Instructional Services	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	0	0.00%	0	0.00%	1	1.10%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	0	0.00%	0	0.00%	0	0.00%
K: Had supervised breaks	0	0.00%	1	0.51%	0	0.00%
L: Most beneficial time of day	0	0.00%	1	0.51%	0	0.00%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	0	0.00%	1	0.51%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in Section 504 Plan	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in IEP	0	0.00%	1	0.51%	0	0.00%
Any Accom., Mod., or EL Variation	0	0.00%	1	0.51%	1	1.10%
No Accom., Mod., or EL Variation	214	100.00%	194	99.49%	90	98.90%

Accommodation Summary for Reading/Language Arts Grades 2, 3, and 4³

EL Program: None (EL only)	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	0	0.00%	0	0.00%	0	0.00%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	0	0.00%	0	0.00%	0	0.00%
K: Had supervised breaks	0	0.00%	0	0.00%	0	0.00%
L: Most beneficial time of day	0	0.00%	0	0.00%	0	0.00%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	0	0.00%	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in Section 504 Plan	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in IEP	0	0.00%	0	0.00%	0	0.00%
Any Accom., Mod., or EL Variation	0	0.00%	0	0.00%	0	0.00%
No Accom., Mod., or EL Variation	164	100.00%	122	100.00%	63	100.00%

Table 6.C.2 Accommodation Summary for Mathematics Grades 2, 3, and 4

Accommodation Summary for Mathematics Grades 2, 3, and 4						
All Students Tested	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	2	0.01%	0	0.00%	3	0.04%
C: Dictated responses to a scribe	3	0.02%	2	0.02%	0	0.00%
F: Used non-interfering assistive device	1	0.01%	1	0.01%	0	0.00%
G: Used Braille Test	1	0.01%	0	0.00%	0	0.00%
H: Used Large Print Test	1	0.01%	3	0.02%	0	0.00%
J: Tested over more than one day	19	0.10%	8	0.06%	3	0.04%
K: Had supervised breaks	36	0.19%	41	0.31%	19	0.24%
L: Most beneficial time of day	14	0.07%	20	0.15%	14	0.18%
M: Administered at home or in a hospital	3	0.02%	0	0.00%	0	0.00%
N: Used a dictionary	1	0.01%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	27	0.14%	40	0.30%	20	0.25%
Q: Used a calculator	1	0.01%	1	0.01%	1	0.01%
R: Used an arithmetic table	1	0.01%	0	0.00%	0	0.00%
S: Used math manipulatives	1	0.01%	0	0.00%	0	0.00%
V: Used interfering assistive device	1	0.01%	0	0.00%	0	0.00%
W: Used an unlisted modification	1	0.01%	3	0.02%	0	0.00%
X: Used an unlisted accommodation	13	0.07%	4	0.03%	8	0.10%
Accom. or Mod. is in Section 504 Plan	3	0.02%	4	0.03%	1	0.01%
Accom. or Mod. is in IEP	54	0.29%	66	0.50%	37	0.46%
<i>Any</i> Accom., Mod., or EL Variation	72	0.38%	78	0.59%	44	0.55%
<i>No</i> Accom., Mod., or EL Variation	18,756	99.62%	13,089	99.41%	7,930	99.45%

Target Students Tested	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	1	0.01%	0	0.00%	0	0.00%
C: Dictated responses to a scribe	2	0.02%	2	0.02%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	1	0.01%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	11	0.09%	3	0.03%	2	0.04%
K: Had supervised breaks	25	0.20%	23	0.26%	8	0.15%
L: Most beneficial time of day	7	0.06%	13	0.14%	10	0.19%
M: Administered at home or in a hospital	2	0.02%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	19	0.15%	22	0.24%	13	0.25%
Q: Used a calculator	0	0.00%	1	0.01%	1	0.02%
R: Used an arithmetic table	0	0.00%	0	0.00%	0	0.00%
S: Used math manipulatives	0	0.00%	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	2	0.02%	0	0.00%
X: Used an unlisted accommodation	7	0.06%	1	0.01%	5	0.09%
Accom. or Mod. is in Section 504 Plan	2	0.02%	2	0.02%	1	0.02%
Accom. or Mod. is in IEP	37	0.29%	37	0.41%	21	0.40%
<i>Any</i> Accom., Mod., or EL Variation	49	0.39%	43	0.48%	25	0.47%
<i>No</i> Accom., Mod., or EL Variation	12,517	99.61%	8,968	99.52%	5,281	99.53%

Accommodation Summary for Mathematics Grades 2, 3, and 4

Non-Target (Optional) Students Tested	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
	B: Marked in test booklet	1	0.02%	0	0.00%	3
C: Dictated responses to a scribe	1	0.02%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	1	0.02%	0	0.00%	0	0.00%
G: Used Braille Test	1	0.02%	0	0.00%	0	0.00%
H: Used Large Print Test	1	0.02%	3	0.07%	0	0.00%
J: Tested over more than one day	8	0.13%	5	0.12%	1	0.04%
K: Had supervised breaks	11	0.18%	18	0.43%	11	0.41%
L: Most beneficial time of day	7	0.11%	7	0.17%	4	0.15%
M: Administered at home or in a hospital	1	0.02%	0	0.00%	0	0.00%
N: Used a dictionary	1	0.02%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	8	0.13%	18	0.43%	7	0.26%
Q: Used a calculator	1	0.02%	0	0.00%	0	0.00%
R: Used an arithmetic table	1	0.02%	0	0.00%	0	0.00%
S: Used math manipulatives	1	0.02%	0	0.00%	0	0.00%
V: Used interfering assistive device	1	0.02%	0	0.00%	0	0.00%
W: Used an unlisted modification	1	0.02%	1	0.02%	0	0.00%
X: Used an unlisted accommodation	6	0.10%	3	0.07%	3	0.11%
Accom. or Mod. is in Section 504 Plan	1	0.02%	2	0.05%	0	0.00%
Accom. or Mod. is in IEP	17	0.27%	29	0.70%	16	0.60%
Any Accom., Mod., or EL Variation	23	0.37%	35	0.84%	19	0.71%
No Accom., Mod., or EL Variation	6,239	99.63%	4,121	99.16%	2,649	99.29%

Students Not in Special Education	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
	B: Marked in test booklet	1	0.01%	0	0.00%	3
C: Dictated responses to a scribe	3	0.02%	2	0.02%	0	0.00%
F: Used non-interfering assistive device	1	0.01%	1	0.01%	0	0.00%
G: Used Braille Test	1	0.01%	0	0.00%	0	0.00%
H: Used Large Print Test	1	0.01%	1	0.01%	0	0.00%
J: Tested over more than one day	3	0.02%	1	0.01%	0	0.00%
K: Had supervised breaks	8	0.04%	10	0.08%	5	0.07%
L: Most beneficial time of day	4	0.02%	10	0.08%	5	0.07%
M: Administered at home or in a hospital	3	0.02%	0	0.00%	0	0.00%
N: Used a dictionary	1	0.01%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	7	0.04%	11	0.09%	6	0.08%
Q: Used a calculator	1	0.01%	0	0.00%	1	0.01%
R: Used an arithmetic table	1	0.01%	0	0.00%	0	0.00%
S: Used math manipulatives	1	0.01%	0	0.00%	0	0.00%
V: Used interfering assistive device	1	0.01%	0	0.00%	0	0.00%
W: Used an unlisted modification	1	0.01%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	4	0.02%	2	0.02%	0	0.00%
Accom. or Mod. is in Section 504 Plan	3	0.02%	4	0.03%	1	0.01%
Accom. or Mod. is in IEP	10	0.06%	12	0.10%	5	0.07%
Any Accom., Mod., or EL Variation	18	0.10%	20	0.16%	11	0.14%
No Accom., Mod., or EL Variation	17,981	99.90%	12,509	99.84%	7,641	99.86%

Accommodation Summary for Mathematics Grades 2, 3, and 4

Students in Special Education	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	1	0.12%	0	0.00%	0	0.00%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	1	0.16%	0	0.00%
J: Tested over more than one day	16	1.95%	6	0.96%	3	0.97%
K: Had supervised breaks	28	3.41%	30	4.82%	10	3.23%
L: Most beneficial time of day	10	1.22%	9	1.44%	5	1.61%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	20	2.44%	29	4.65%	10	3.23%
Q: Used a calculator	0	0.00%	1	0.16%	0	0.00%
R: Used an arithmetic table	0	0.00%	0	0.00%	0	0.00%
S: Used math manipulatives	0	0.00%	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	2	0.32%	0	0.00%
X: Used an unlisted accommodation	9	1.10%	2	0.32%	8	2.58%
Accom. or Mod. is in Section 504 Plan	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in IEP	44	5.37%	51	8.19%	28	9.03%
Any Accom., Mod., or EL Variation	54	6.59%	55	8.83%	29	9.35%
No Accom., Mod., or EL Variation	766	93.41%	568	91.17%	281	90.65%

Students in U.S. Schools < 12 Months	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	0	0.00%	0	0.00%	0	0.00%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	0	0.00%	0	0.00%	0	0.00%
K: Had supervised breaks	0	0.00%	5	0.19%	1	0.04%
L: Most beneficial time of day	0	0.00%	4	0.15%	2	0.08%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	1	0.03%	3	0.11%	2	0.08%
Q: Used a calculator	0	0.00%	0	0.00%	1	0.04%
R: Used an arithmetic table	0	0.00%	0	0.00%	0	0.00%
S: Used math manipulatives	0	0.00%	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	0	0.00%	1	0.04%	0	0.00%
Accom. or Mod. is in Section 504 Plan	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in IEP	0	0.00%	5	0.19%	2	0.08%
Any Accom., Mod., or EL Variation	1	0.03%	7	0.26%	3	0.12%
No Accom., Mod., or EL Variation	2,857	99.97%	2,640	99.74%	2,445	99.88%

Accommodation Summary for Mathematics Grades 2, 3, and 4

Students in U.S. Schools \geq 12 Months	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
	B: Marked in test booklet	2	0.01%	0	0.00%	3
C: Dictated responses to a scribe	3	0.02%	2	0.02%	0	0.00%
F: Used non-interfering assistive device	1	0.01%	1	0.01%	0	0.00%
G: Used Braille Test	1	0.01%	0	0.00%	0	0.00%
H: Used Large Print Test	1	0.01%	3	0.03%	0	0.00%
J: Tested over more than one day	19	0.12%	8	0.08%	3	0.05%
K: Had supervised breaks	36	0.23%	36	0.34%	18	0.33%
L: Most beneficial time of day	14	0.09%	16	0.15%	12	0.22%
M: Administered at home or in a hospital	3	0.02%	0	0.00%	0	0.00%
N: Used a dictionary	1	0.01%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	26	0.16%	37	0.35%	18	0.33%
Q: Used a calculator	1	0.01%	1	0.01%	0	0.00%
R: Used an arithmetic table	1	0.01%	0	0.00%	0	0.00%
S: Used math manipulatives	1	0.01%	0	0.00%	0	0.00%
V: Used interfering assistive device	1	0.01%	0	0.00%	0	0.00%
W: Used an unlisted modification	1	0.01%	3	0.03%	0	0.00%
X: Used an unlisted accommodation	13	0.08%	3	0.03%	8	0.14%
Accom. or Mod. is in Section 504 Plan	3	0.02%	4	0.04%	1	0.02%
Accom. or Mod. is in IEP	54	0.34%	61	0.58%	35	0.63%
Any Accom., Mod., or EL Variation	71	0.44%	71	0.67%	41	0.74%
No Accom., Mod., or EL Variation	15,899	99.56%	10,449	99.33%	5,485	99.26%

EL Program: EL in ELD	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
	B: Marked in test booklet	0	0.00%	0	0.00%	0
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	0	0.00%	0	0.00%	0	0.00%
K: Had supervised breaks	0	0.00%	0	0.00%	0	0.00%
L: Most beneficial time of day	0	0.00%	0	0.00%	1	0.20%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	1	0.13%	2	0.33%	0	0.00%
Q: Used a calculator	0	0.00%	0	0.00%	0	0.00%
R: Used an arithmetic table	0	0.00%	0	0.00%	0	0.00%
S: Used math manipulatives	0	0.00%	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	3	0.38%	3	0.49%	3	0.59%
Accom. or Mod. is in Section 504 Plan	0	0.00%	1	0.16%	0	0.00%
Accom. or Mod. is in IEP	4	0.51%	4	0.66%	3	0.59%
Any Accom., Mod., or EL Variation	4	0.51%	5	0.82%	4	0.78%
No Accom., Mod., or EL Variation	787	99.49%	604	99.18%	508	99.22%

Accommodation Summary for Mathematics Grades 2, 3, and 4

EL Program: EL in ELD and SDAIE	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	0	0.00%	0	0.00%	0	0.00%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	3	0.21%	1	0.07%	1	0.08%
K: Had supervised breaks	3	0.21%	7	0.52%	4	0.33%
L: Most beneficial time of day	1	0.07%	2	0.15%	0	0.00%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	1	0.07%	2	0.15%	2	0.16%
Q: Used a calculator	0	0.00%	0	0.00%	1	0.08%
R: Used an arithmetic table	0	0.00%	0	0.00%	0	0.00%
S: Used math manipulatives	0	0.00%	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	0	0.00%	1	0.07%	0	0.00%
Accom. or Mod. is in Section 504 Plan	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in IEP	3	0.21%	7	0.52%	5	0.41%
Any Accom., Mod., or EL Variation	4	0.28%	9	0.67%	5	0.41%
No Accom., Mod., or EL Variation	1,416	99.72%	1,341	99.33%	1,214	99.59%

EL Program: EL in ELD and SDAIE with Primary Language Support	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	0	0.00%	0	0.00%	0	0.00%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	1	0.04%	0	0.00%	0	0.00%
K: Had supervised breaks	2	0.08%	2	0.12%	3	0.21%
L: Most beneficial time of day	1	0.04%	1	0.06%	0	0.00%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	0	0.00%	2	0.12%	1	0.07%
Q: Used a calculator	0	0.00%	0	0.00%	0	0.00%
R: Used an arithmetic table	0	0.00%	0	0.00%	0	0.00%
S: Used math manipulatives	0	0.00%	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in Section 504 Plan	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in IEP	2	0.08%	2	0.12%	4	0.28%
Any Accom., Mod., or EL Variation	2	0.08%	3	0.17%	4	0.28%
No Accom., Mod., or EL Variation	2,390	99.92%	1,714	99.83%	1,410	99.72%

Accommodation Summary for Mathematics Grades 2, 3, and 4

EL Program: EL in ELD and Academic Subjects through Primary Language	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	2	0.02%	0	0.00%	2	0.05%
C: Dictated responses to a scribe	3	0.02%	2	0.02%	0	0.00%
F: Used non-interfering assistive device	1	0.01%	1	0.01%	0	0.00%
G: Used Braille Test	1	0.01%	0	0.00%	0	0.00%
H: Used Large Print Test	1	0.01%	2	0.02%	0	0.00%
J: Tested over more than one day	15	0.11%	6	0.07%	2	0.05%
K: Had supervised breaks	30	0.23%	29	0.33%	8	0.18%
L: Most beneficial time of day	11	0.08%	15	0.17%	9	0.20%
M: Administered at home or in a hospital	3	0.02%	0	0.00%	0	0.00%
N: Used a dictionary	1	0.01%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	24	0.18%	33	0.38%	13	0.29%
Q: Used a calculator	1	0.01%	1	0.01%	0	0.00%
R: Used an arithmetic table	1	0.01%	0	0.00%	0	0.00%
S: Used math manipulatives	1	0.01%	0	0.00%	0	0.00%
V: Used interfering assistive device	1	0.01%	0	0.00%	0	0.00%
W: Used an unlisted modification	1	0.01%	2	0.02%	0	0.00%
X: Used an unlisted accommodation	10	0.08%	0	0.00%	5	0.11%
Accom. or Mod. is in Section 504 Plan	3	0.02%	3	0.03%	1	0.02%
Accom. or Mod. is in IEP	44	0.34%	48	0.55%	21	0.48%
Any Accom., Mod., or EL Variation	61	0.47%	56	0.65%	26	0.59%
No Accom., Mod., or EL Variation	13,034	99.53%	8,623	99.35%	4,394	99.41%

EL Program: Other EL Instructional Services	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	0	0.00%	0	0.00%	1	1.10%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	0	0.00%	0	0.00%	0	0.00%
K: Had supervised breaks	0	0.00%	1	0.53%	0	0.00%
L: Most beneficial time of day	0	0.00%	1	0.53%	0	0.00%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	0	0.00%	1	0.53%	0	0.00%
Q: Used a calculator	0	0.00%	0	0.00%	0	0.00%
R: Used an arithmetic table	0	0.00%	0	0.00%	0	0.00%
S: Used math manipulatives	0	0.00%	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in Section 504 Plan	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in IEP	0	0.00%	1	0.53%	0	0.00%
Any Accom., Mod., or EL Variation	0	0.00%	1	0.53%	1	1.10%
No Accom., Mod., or EL Variation	210	100.00%	188	99.47%	90	98.90%

Accommodation Summary for Mathematics Grades 2, 3, and 4

EL Program: None (EL only)	Grade 2		Grade 3		Grade 4	
	Count	Pct. of Total	Count	Pct. of Total	Count	Pct. of Total
B: Marked in test booklet	0	0.00%	0	0.00%	0	0.00%
C: Dictated responses to a scribe	0	0.00%	0	0.00%	0	0.00%
F: Used non-interfering assistive device	0	0.00%	0	0.00%	0	0.00%
G: Used Braille Test	0	0.00%	0	0.00%	0	0.00%
H: Used Large Print Test	0	0.00%	0	0.00%	0	0.00%
J: Tested over more than one day	0	0.00%	0	0.00%	0	0.00%
K: Had supervised breaks	0	0.00%	0	0.00%	0	0.00%
L: Most beneficial time of day	0	0.00%	0	0.00%	0	0.00%
M: Administered at home or in a hospital	0	0.00%	0	0.00%	0	0.00%
N: Used a dictionary	0	0.00%	0	0.00%	0	0.00%
P: Examiner read test questions aloud	0	0.00%	0	0.00%	0	0.00%
Q: Used a calculator	0	0.00%	0	0.00%	0	0.00%
R: Used an arithmetic table	0	0.00%	0	0.00%	0	0.00%
S: Used math manipulatives	0	0.00%	0	0.00%	0	0.00%
V: Used interfering assistive device	0	0.00%	0	0.00%	0	0.00%
W: Used an unlisted modification	0	0.00%	0	0.00%	0	0.00%
X: Used an unlisted accommodation	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in Section 504 Plan	0	0.00%	0	0.00%	0	0.00%
Accom. or Mod. is in IEP	0	0.00%	0	0.00%	0	0.00%
Any Accom., Mod., or EL Variation	0	0.00%	0	0.00%	0	0.00%
No Accom., Mod., or EL Variation	162	100.00%	122	100.00%	63	100.00%

Appendix 6.D — DIF Tables

Table 6.D.1 Field-test Items Exhibiting Significant DIF

Test	Item Number	Form	Item Seq. No.	Male-Female
MTHMGR03	STM11778	7	66	C+

Table 6.D.2 Male-Female DIF Classifications for Operational Items

DIF Category	Reading/Language Arts						Mathematics					
	Grade 2		Grade 3		Grade 4		Grade 2		Grade 3		Grade 4	
	N	Pct.	N	Pct.	N	Pct.	N	Pct.	N	Pct.	N	Pct.
C-	0	0	0	0	0	0	0	0	0	0	0	0
B-	0	0	0	0	0	0	1	2	3	5	1	2
A-	34	52	31	48	36	48	28	43	26	40	32	49
A+	30	46	34	52	39	52	34	52	36	55	31	48
B+	1	2	0	0	0	0	2	3	0	0	1	2
C+	0	0	0	0	0	0	0	0	0	0	0	0
Small N	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	65	100	65	100	75	100	65	100	65	100	65	100

Table 6.D.3 Male-Female DIF Classifications for Field-test Items

DIF Category	Reading/Language Arts						Mathematics					
	Grade 2		Grade 3		Grade 4		Grade 2		Grade 3		Grade 4	
	N	Pct.	N	Pct.	N	Pct.	N	Pct.	N	Pct.	N	Pct.
C-	0	0	0	0	0	0	0	0	0	0	0	0
B-	0	0	1	1	0	0	2	3	0	0	2	3
A-	40	56	35	49	11	15	30	43	37	51	9	13
A+	31	44	34	47	13	18	36	51	29	40	13	18
B+	0	0	2	3	0	0	2	3	5	7	0	0
C+	0	0	0	0	0	0	0	0	1	1	0	0
Small N	0	0	0	0	48	67	0	0	0	0	48	67
TOTAL ⁵	71	100	72	100	72	100	70	100	72	100	72	100

⁵ A couple of field-test items appeared on more than one form, so the total might not equal 72.

Chapter 7: Item Quality

This chapter summarizes the item-level statistics obtained for the California Standards-based Tests in Spanish program administered during spring of 2007. This includes STS for RLA and STS for Mathematics items for grades two to four. Each STS was composed of dichotomously scored multiple-choice (MC) items.

The STS tests also included blocks of six field-test items that were not included in the operational test scores. Different sets of items were presented in each form (version)⁶ for the various STS tests. In all STS tests for RLA and math, the items were field-tested in twelve forms.

The item-level IRT information presented in this chapter is based on the complete STS data set. Table 7.1 summarizes information about the test forms⁷ and students included in the item analyses, including the numbers of test forms, operational items, field-test items, and the approximate number of students taking operational and field-test items.

Table 7.1 Summary of Items and Forms Presented in the 2007 STS

Subject	STS	Operational		Field Test		
		# Items	# Examinees	# Forms	# Items	# Examinees per form
Reading/Language Arts	2	65	12,486	12	71	694 – 1,939
	3	65	8,952	12	72	720 – 781
	4	75	5,291	12	72	371 – 422
Mathematics	2	65	12,469	12	70	697 – 1,936
	3	65	8,926	12	72	715 – 778
	4	65	5,283	12	72	371 – 420

The statistics presented in this chapter are divided into two sections as follows:

1. Summaries of classical item-level analyses, including item proportion correct (p-value), point-biserial correlations (Pt-Rbis) for each operational item, and summaries of overall p-value and point-biserial correlation statistics across operational items. These statistics are presented in Appendix 7.A.
2. Summaries of Rasch model item difficulty statistics (b-values) for operational and field-test items, and summaries of item classifications based on the fit of the Rasch model to the data, reported in letter categories of A, B, C, D, and F (IRT flag) for operational and field-test items. These statistics are presented in Appendix 7.B.

DIF analyses were also performed on all operational items and all field-test items for which sufficient student samples were available. Those results are presented in Appendix 6.D of Chapter 6.

Item Analyses

This section describes the overall and item-by-item proportion correct indices as well as the point-biserial correlation indices for the operational items. The point-biserial correlation is a special case of the Pearson product-moment correlation used to measure the relationship between two variables, one dichotomous 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:

⁶ A version of a test is one that has the same operational form of the test with different field-test item sets. These are considered different *forms* of the same test.

⁷ A form was counted as a field-test form if it contains one or more field-test items.

$$R_{it} = \frac{Cov(i,t)}{\sigma_{xi}\sigma_t} \quad (7.1)$$

where,

$Cov(i,t)$ is the Covariance between an item i and total score t

σ_{xi} is the standard deviation for an item i

σ_t is the standard deviation for t

Table 7.2 presents summary indices for each operational test. Both the mean and median are provided. Detailed tables are provided in Appendix 7.A.

The data in Table 7.2 indicate that all STS tests have average p-values between 0.5 and 0.7⁸. The tests that were easiest were mathematics administered at grades 2 (mean p-value = 0.69) and grade 3 (mean p-value = 0.67). The STS tests that were relatively difficult for these students (mean p-value ≤ 0.60) were RLA grades 3 and 4.

The average item-total correlations indicated levels of item discrimination that were similar to expectations. This index was greater than or equal to 0.38 for all STS tests. The math tests exhibited higher item-total correlations, with mean values all at or above 0.40.

The item-by-item values for the indices are presented in Table 7.A.1 which starts on page 91.

Table 7.2 Average and Median Proportion Correct and Point-Biserial

Subject	STS	Number of items	Number of Examinees	Mean		Median	
				p-value	Pt-Rbis	p-value	Pt-Rbis
<i>Reading/Language Arts</i>	2	65	12,486	0.63	0.41	0.65	0.44
	3	65	8,952	0.57	0.37	0.56	0.38
	4	75	5,291	0.54	0.38	0.54	0.40
<i>Mathematics</i>	2	65	12,469	0.69	0.40	0.73	0.42
	3	65	8,926	0.67	0.43	0.68	0.44
	4	65	5,283	0.61	0.42	0.62	0.44

IRT Analyses

The results of the IRT analyses are presented in Appendix 7.B, which starts on page 93. Table 7.B.1 through Table 7.B.8 present summary univariate statistics (mean, standard deviation, minimum, and maximum) and distributions for the IRT b-values. These statistics are listed for the operational test, by cluster scores, and for the field-test items. Table 7.B.9 through Table 7.B.10 present the results of the IRT model-data fit analyses for the STS.

Summaries of IRT b-values

The summary of IRT b-values for the operational and field test items are presented in Table 7.B.1 through Table 7.B.6 in Appendix 7.B.

Table 7.B.7 to Table 7.B.8 depict a more detailed presentation of IRT b-values for the operational and field-test items. The tables show the distribution of items at 16 IRT b-value intervals, where the b-values ranged from “less than -3.5” to “greater than or equal to 3.5.” These distributions indicated that most of the items had difficulty levels in the range of -2.0 and 2.0 for all the STS tests.

⁸ The average p-value can differ from one test to another for many reasons, including the perspective of the Test Development committee, the available items in the assembly pool, changes in the population taking the test from the population used to develop the target, and discrepancies between the target and the assembled test.

The overall difficulty level of the field-test items was higher than the operational items for all STS tests. The average difficulty level for the operational items ranged from -0.96 for grade 2 math to -0.18 for grade 4 RLA. The index of average difficulty for the field-test items ranged from -0.89 for grade 2 math to 0.57 for grade 4 RLA.

IRT Model-Data Fit Analyses

Because the Rasch model will be used in equating and scaling the STS tests, an important part of IRT item analyses is the assessment of model-data fit. ETS statisticians classified operational and field-test items for the STS into discrete categories based on an evaluation of how well each item was fit by the Rasch model. The flagging procedure has categories of A, B, C, D, and F that are assigned based on an evaluation of graphical model-data fit information. Descriptors for each category are provided below. As an illustration, the IRT item characteristic curves and empirical data (item-ability regressions) for five CST items field-tested in 2005 are shown in Figure 7.1. These five items represent the various rating categories. The item number in the calibration and ETS identification number for each item (“accession number”) are listed next to each item as well as the corresponding rating categories.

Flag A (CST Item 236, CSV23487)

- Good fit of theoretical curve to empirical data along the entire ability range, may have some small divergence at the extremes
- Small Chi-square value relative to the other items in the calibration with similar sample sizes

Flag B (CST Item 061, CSV22589)

- Theoretical curve within error range across most of ability range, may have some small divergence at the extremes
- Acceptable Chi-square value relative to the other items in the calibration with similar sample sizes

Flag C (CST Item 165, CSV20282)

- Theoretical curve within error range at some regions and slightly outside of error range at remaining regions of ability range
- Moderate Chi-square value relative to the other items in the calibration with similar sample sizes
- This category often applies to items that appear to be functioning well, but that are not well fit by the Rasch model

Flag D (CST Item 113, CSV20317)

- Theoretical curve outside of error range at some regions across ability range
- Large Chi-square value relative to the other items in the calibration with similar sample sizes

Flag F (CST Item 184, CSV20311)

- Theoretical curve outside of error range at most regions across ability range
- Probability of answering item correctly may be higher at lower ability than higher ability (U-shaped empirical curve)
- Very large Chi-square value relative to the other items with similar sample sizes and classical item statistics tend also to be very poor.

In general, items with flagging categories of A, B, or C are all considered acceptable. Ratings of D are considered questionable—test developers are asked to avoid these items if possible and to

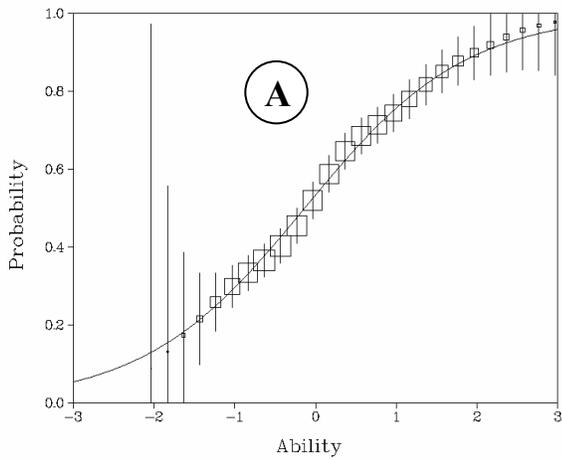
carefully review them if they must be used. Test developers are instructed to avoid using items rated F for operational test assembly without a review by a psychometrician.

The results of the IRT model data fit classifications are presented in Table 7.B.9 for operational items and in Table 7.B.10 for the field-test items. In general, very few operational items were rated as F items. For RLA, 8 out of 205 operational items were flagged as F items. For math, only 2 out of 130 items were flagged as F items.

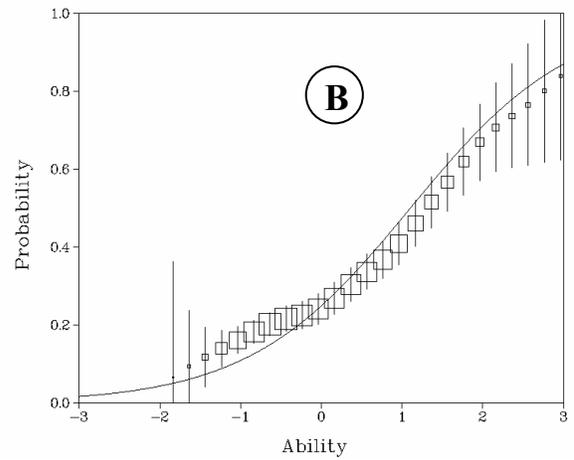
Summary of Item-Level Analyses

The item-level analyses of the 2007 STS tests presented in this chapter included classical item analyses and IRT analyses. Overall, these analyses indicate that the STS tests administered in 2007 meet the technical criteria established in professional standards for statewide standard tests, and that the items field-tested as part of the 2007 administration have statistical characteristics that are appropriate for use in future administrations.

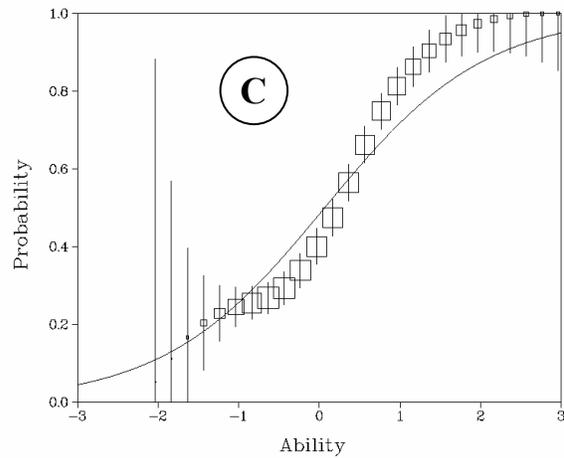
Figure 7.1 Items from the 2005 CST History–Social Science Grade 10 Field-test Calibration



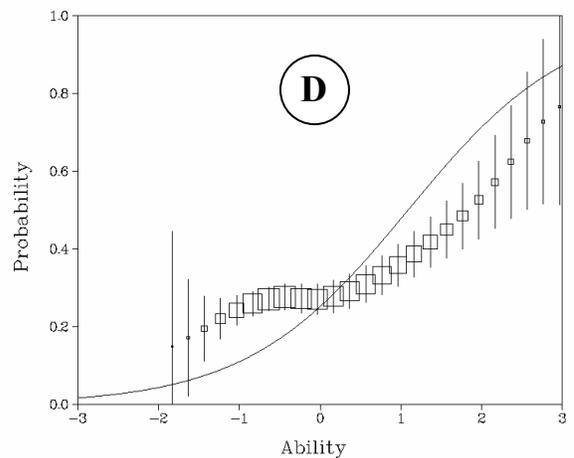
Version 30, Seq 29 (#236) CSV23487 4-Choice P+=0.563
 $a=0.588$ F, $b=-0.135$, $c=0.000$ F, CHI=5.41, N=5,912



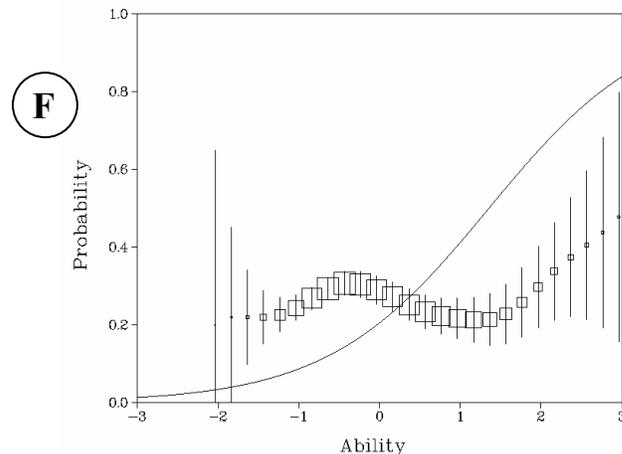
Version 1, Seq 28 (#61) CSV22589 4 Choice P+=0.307
 $a=0.588$ F, $b=1.104$, $c=0.000$ F, CHI=66.70, N=6,348



Version 18, Seq 30 (#165) CSV20282 4-Choice P+=0.523
 $a=0.588$ F, $b=0.066$, $c=0.000$ F, CHI=208.99, N=6,183



Version 9, Seq 32 (#113) CSV20317 4-Choice P+=0.314
 $a=0.588$ F, $b=1.089$, $c=0.000$ F, CHI=361.31, N=6,047



Version 21, Seq 31 (#184) CSV20311 4-Choice P+=0.263
 $a=0.588$ F, $b=1.356$, $c=0.000$ F, CHI=1027.57, N=6,277

References

Holland, P.W. and Thayer, D.T. "An alternative definition of the ETS delta scale of item difficulty." RR-85-43, 1985.

Hambleton, R.K. and Swaminathan, H. *Item Response Theory: Principles and Applications*. Boston, MA: Kluwer-Nijhoff, 1985.

Appendix 7.A—Item-by-Item p-value and Point-Biserial Tables

Table 7.A.1 Item-by-item p-value and Point-Biserial

Items	Reading/Language Arts						Mathematics					
	Grade 2		Grade 3		Grade 4		Grade 2		Grade 3		Grade 4	
	p-value	Pt-Rbis	p-value	Pt-Rbis	p-value	Pt-Rbis	p-value	Pt-Rbis	p-value	Pt-Rbis	p-value	Pt-Rbis
1	0.87	0.31	0.88	0.22	0.66	0.46	0.92	0.32	0.67	0.45	0.49	0.39
2	0.33	0.23	0.91	0.27	0.57	0.26	0.78	0.43	0.67	0.43	0.61	0.33
3	0.90	0.35	0.70	0.35	0.41	0.36	0.52	0.49	0.66	0.29	0.77	0.45
4	0.72	0.45	0.28	0.22	0.81	0.36	0.90	0.31	0.81	0.42	0.65	0.33
5	0.63	0.32	0.39	0.22	0.45	0.38	0.84	0.43	0.62	0.34	0.43	0.38
6	0.59	0.31	0.81	0.32	0.68	0.36	0.74	0.46	0.50	0.47	0.74	0.44
7	0.80	0.42	0.56	0.41	0.70	0.34	0.87	0.30	0.77	0.52	0.51	0.55
8	0.75	0.52	0.82	0.39	0.70	0.41	0.77	0.42	0.60	0.53	0.80	0.41
9	0.80	0.50	0.84	0.43	0.63	0.32	0.73	0.50	0.70	0.57	0.67	0.48
10	0.67	0.48	0.71	0.49	0.52	0.42	0.64	0.45	0.74	0.40	0.63	0.44
11	0.67	0.54	0.51	0.32	0.35	0.30	0.73	0.21	0.60	0.56	0.63	0.40
12	0.60	0.50	0.56	0.34	0.33	0.22	0.68	0.42	0.38	0.47	0.86	0.37
13	0.68	0.46	0.34	0.20	0.56	0.44	0.52	0.41	0.74	0.43	0.51	0.48
14	0.65	0.36	0.73	0.38	0.69	0.45	0.42	0.45	0.74	0.40	0.62	0.46
15	0.76	0.38	0.50	0.42	0.79	0.44	0.55	0.44	0.42	0.48	0.60	0.49
16	0.61	0.38	0.31	0.22	0.44	0.41	0.52	0.42	0.57	0.40	0.86	0.37
17	0.73	0.55	0.63	0.51	0.61	0.37	0.77	0.43	0.78	0.42	0.45	0.36
18	0.59	0.32	0.72	0.48	0.45	0.32	0.55	0.33	0.55	0.52	0.60	0.45
19	0.31	0.23	0.58	0.44	0.47	0.45	0.52	0.49	0.57	0.50	0.71	0.45
20	0.36	0.19	0.84	0.29	0.23	0.07	0.78	0.36	0.87	0.44	0.57	0.33
21	0.57	0.38	0.58	0.46	0.34	0.21	0.60	0.39	0.81	0.40	0.69	0.46
22	0.28	0.20	0.75	0.46	0.48	0.34	0.82	0.33	0.57	0.46	0.86	0.35
23	0.38	0.34	0.44	0.36	0.43	0.35	0.80	0.47	0.58	0.45	0.48	0.37
24	0.80	0.40	0.86	0.45	0.60	0.20	0.49	0.50	0.73	0.51	0.64	0.50
25	0.86	0.41	0.56	0.41	0.44	0.39	0.47	0.47	0.82	0.50	0.74	0.50
26	0.86	0.41	0.54	0.38	0.64	0.52	0.59	0.38	0.94	0.34	0.51	0.45
27	0.85	0.38	0.41	0.37	0.59	0.50	0.70	0.46	0.31	0.36	0.46	0.37
28	0.61	0.43	0.92	0.35	0.50	0.39	0.58	0.38	0.73	0.50	0.43	0.45
29	0.60	0.49	0.45	0.46	0.57	0.47	0.64	0.43	0.78	0.49	0.62	0.50
30	0.86	0.50	0.36	0.33	0.45	0.39	0.53	0.54	0.66	0.47	0.48	0.35
31	0.62	0.45	0.35	0.26	0.58	0.54	0.94	0.36	0.82	0.50	0.43	0.32
32	0.58	0.46	0.37	0.17	0.51	0.38	0.62	0.47	0.59	0.37	0.52	0.34
33	0.83	0.51	0.51	0.37	0.67	0.45	0.95	0.24	0.53	0.38	0.52	0.42
34	0.38	0.23	0.67	0.47	0.57	0.36	0.42	0.35	0.78	0.48	0.78	0.49
35	0.77	0.53	0.68	0.48	0.25	0.13	0.38	0.34	0.48	0.53	0.67	0.45
36	0.80	0.50	0.43	0.38	0.20	0.06	0.77	0.44	0.78	0.44	0.78	0.48
37	0.71	0.51	0.35	0.40	0.52	0.38	0.53	0.47	0.72	0.51	0.51	0.41
38	0.75	0.53	0.49	0.39	0.86	0.40	0.89	0.39	0.46	0.41	0.63	0.47
39	0.65	0.48	0.64	0.47	0.54	0.29	0.52	0.42	0.61	0.55	0.66	0.51
40	0.73	0.59	0.59	0.39	0.79	0.48	0.49	0.41	0.73	0.49	0.62	0.40
41	0.59	0.48	0.71	0.47	0.69	0.50	0.73	0.31	0.68	0.50	0.68	0.51
42	0.71	0.53	0.51	0.30	0.51	0.42	0.74	0.48	0.49	0.48	0.69	0.54
43	0.55	0.26	0.45	0.27	0.62	0.42	0.72	0.46	0.35	0.25	0.70	0.53
44	0.38	0.22	0.54	0.48	0.73	0.55	0.63	0.45	0.81	0.44	0.70	0.55
45	0.63	0.54	0.58	0.29	0.39	0.14	0.53	0.24	0.69	0.29	0.61	0.50
46	0.48	0.40	0.58	0.45	0.70	0.47	0.77	0.30	0.61	0.40	0.55	0.51

Items	Reading/Language Arts						Mathematics					
	Grade 2		Grade 3		Grade 4		Grade 2		Grade 3		Grade 4	
	p-value	Pt-Rbis	p-value	Pt-Rbis	p-value	Pt-Rbis	p-value	Pt-Rbis	p-value	Pt-Rbis	p-value	Pt-Rbis
47	0.69	0.38	0.59	0.46	0.73	0.43	0.71	0.31	0.56	0.42	0.59	0.53
48	0.57	0.45	0.53	0.33	0.45	0.32	0.79	0.29	0.64	0.35	0.59	0.57
49	0.34	0.25	0.60	0.43	0.65	0.57	0.56	0.25	0.89	0.44	0.56	0.47
50	0.37	0.17	0.56	0.47	0.58	0.49	0.83	0.33	0.90	0.27	0.62	0.42
51	0.40	0.24	0.45	0.31	0.48	0.37	0.80	0.31	0.58	0.41	0.83	0.37
52	0.39	0.21	0.26	0.14	0.22	0.13	0.91	0.35	0.92	0.36	0.58	0.42
53	0.49	0.38	0.67	0.47	0.46	0.30	0.80	0.39	0.66	0.45	0.48	0.34
54	0.74	0.55	0.63	0.42	0.50	0.44	0.75	0.43	0.82	0.39	0.58	0.36
55	0.40	0.28	0.50	0.33	0.40	0.22	0.86	0.38	0.85	0.40	0.77	0.40
56	0.87	0.45	0.49	0.36	0.63	0.42	0.58	0.50	0.69	0.32	0.67	0.44
57	0.67	0.55	0.56	0.37	0.54	0.45	0.77	0.34	0.49	0.33	0.60	0.35
58	0.59	0.50	0.40	0.27	0.56	0.38	0.87	0.44	0.53	0.52	0.38	0.27
59	0.36	0.19	0.48	0.42	0.36	0.24	0.84	0.48	0.63	0.23	0.65	0.28
60	0.67	0.53	0.53	0.32	0.48	0.43	0.62	0.45	0.82	0.27	0.53	0.29
61	0.64	0.50	0.47	0.50	0.59	0.57	0.76	0.39	0.83	0.46	0.66	0.39
62	0.69	0.60	0.50	0.53	0.52	0.47	0.80	0.43	0.82	0.52	0.70	0.46
63	0.74	0.44	0.63	0.51	0.58	0.48	0.83	0.32	0.61	0.49	0.36	0.29
64	0.69	0.56	0.85	0.46	0.44	0.15	0.77	0.48	0.78	0.36	0.57	0.27
65	0.65	0.54	0.44	0.21	0.47	0.21	0.40	0.46	0.76	0.47	0.54	0.44
66					0.39	0.42						
67					0.63	0.50						
68					0.66	0.44						
69					0.72	0.54						
70					0.42	0.25						
71					0.59	0.45						
72					0.69	0.50						
73					0.39	0.25						
74					0.63	0.48						
75					0.47	0.51						

Appendix 7.B—IRT Tables

Table 7.B.1 IRT b-values for Reading/Language Arts Grade 2

Content Area	Number of Items	Mean	Standard Deviation	Minimum	Maximum
Word Analysis and Vocabulary Development	22	-0.92	0.97	-2.50	1.24
Reading Comprehension	15	-0.46	0.76	-1.61	1.04
Literary Response and Analysis	6	-1.22	0.87	-2.09	-0.41
Written Conventions	14	-0.68	0.66	-2.13	0.63
Writing Strategies	8	0.50	0.41	-0.34	0.88
All Operational Items	65	-0.62	0.91	-2.50	1.24
Field-test Items	71	-0.03	0.88	-2.35	1.83

Table 7.B.2 IRT b-values for Reading/Language Arts Grade 3

Content Area	Number of Items	Mean	Standard Deviation	Minimum	Maximum
Word Analysis and Vocabulary Development	20	-0.98	0.98	-2.62	0.35
Reading Comprehension	15	-0.00	0.67	-1.14	0.97
Literary Response and Analysis	8	-0.25	1.18	-1.89	1.12
Written Conventions	13	-0.16	0.32	-0.77	0.27
Writing Strategies	9	0.05	0.55	-0.61	1.22
All Operational Items	65	-0.36	0.88	-2.62	1.22
Field-test Items	72	0.23	0.77	-2.19	1.50

Table 7.B.3 IRT b-values for Reading/Language Arts Grade 4

Content Area	Number of Items	Mean	Standard Deviation	Minimum	Maximum
Word Analysis and Vocabulary Development	18	-0.47	0.61	-1.67	0.54
Reading Comprehension	15	0.03	0.63	-0.94	1.40
Literary Response and Analysis	9	0.36	0.78	-0.74	1.61
Written Conventions	18	-0.42	0.67	-2.01	0.53
Writing Strategies	15	-0.06	0.73	-1.12	1.46
All Operational Items	75	-0.18	0.71	-2.01	1.61
Field-test Items	72	0.57	0.54	-0.95	1.92

Table 7.B.4 IRT b-values for Mathematics Grade 2

Content Area	Number of Items	Mean	Standard Deviation	Minimum	Maximum
Place Value, Addition, and Subtraction	15	-0.86	0.88	-2.70	0.29
Multiplication, Division, and Fractions	23	-0.88	1.15	-3.35	0.76
Algebra and Functions	6	-0.58	0.60	-1.13	0.20
Measurement and Geometry	14	-1.30	0.75	-2.52	-0.00
Statistics, Data Analysis, and Probability	7	-1.08	0.87	-1.84	0.61
All Operational Items	65	-0.96	0.94	-3.35	0.76
Field-test Items	70	-0.89	0.99	-2.96	1.58

Table 7.B.5 IRT b-values for Mathematics Grade 3

Content Area	Number of Items	Mean	Standard Deviation	Minimum	Maximum
Place Value, Fractions, and Decimals	16	-0.69	0.88	-2.18	1.16
Addition, Subtraction, Multiplication, Division	16	-0.90	0.88	-3.15	0.53
Algebra and Functions	12	-0.51	0.84	-1.64	0.96
Measurement and Geometry	16	-1.07	0.94	-2.78	0.18
Statistics, Data Analysis, and Probability	5	-1.32	0.55	-1.77	-0.41
All Operational Items	65	-0.85	0.88	-3.15	1.16
Field-test Items	72	0.05	1.06	-2.04	2.46

Table 7.B.6 IRT b-values for Mathematics Grade 4

Content Area	Number of Items	Mean	Standard Deviation	Minimum	Maximum
Decimals, Fractions, and Negative Numbers	17	-0.76	0.74	-2.09	0.33
Operations and Factoring	14	-0.27	0.69	-1.60	0.45
Algebra and Functions	18	-0.59	0.44	-1.45	0.01
Measurement and Geometry	12	-0.51	0.66	-1.78	0.69
Statistics, Data Analysis, and Probability	4	-0.13	0.71	-0.93	0.81
All Operational Items	65	-0.52	0.65	-2.09	0.81
Field-test Items	72	0.34	0.79	-1.44	1.91

Table 7.B.7 Distribution of IRT b-values for Operational Items

IRT b-value	Reading/Language Arts			Mathematics		
	Grade 2	Grade 3	Grade 4	Grade 2	Grade 3	Grade 4
>= 3.5	0	0	0	0	0	0
3.0 - < 3.5	0	0	0	0	0	0
2.5 - < 3.0	0	0	0	0	0	0
2.0 - < 2.5	0	0	0	0	0	0
1.5 - < 2.0	0	0	1	0	0	0
1.0 - < 1.5	2	2	3	0	1	0
0.5 - < 1.0	11	8	7	3	3	2
0.0 - < 0.5	2	13	19	9	6	13
-0.5 - < 0.0	12	19	19	11	15	16
-1.0 - < -0.5	15	10	18	6	11	21
-1.5 - < -1.0	10	3	5	17	13	7
-2.0 - < -1.5	6	5	1	9	10	2
-2.5 - < -2.0	6	2	1	5	3	3
-3.0 - < -2.5	0	2	0	2	1	0
-3.5 - < -3.0	0	0	0	2	1	0
< -3.5	1	1	1	1	1	1
TOTAL	65	65	75	65	65	65

Table 7.B.8 Distribution of IRT b-values for Field-test Items

IRT b-value	Reading/Language Arts			Mathematics		
	Grade 2	Grade 3	Grade 4	Grade 2	Grade 3	Grade 4
≥ 3.5	0	0	0	0	0	0
$3.0 < 3.5$	0	0	0	0	0	0
$2.5 < 3.0$	0	0	0	0	0	0
$2.0 < 2.5$	0	0	0	0	1	0
$1.5 < 2.0$	2	1	3	1	5	5
$1.0 < 1.5$	5	12	11	1	8	10
$0.5 < 1.0$	14	14	23	5	11	12
$0.0 < 0.5$	17	22	23	6	17	19
$-0.5 < 0.0$	12	12	9	9	8	14
$-1.0 < -0.5$	10	6	2	19	8	8
$-1.5 < -1.0$	7	1	0	9	5	3
$-2.0 < -1.5$	0	1	0	11	6	0
$-2.5 < -2.0$	3	2	0	5	2	0
$-3.0 < -2.5$	0	0	0	3	0	0
$-3.5 < -3.0$	0	0	0	0	0	0
< -3.5	1	1	1	1	1	1
TOTAL	71	72	72	70	72	72

Table 7.B.9 IRT Model Data Fit Distribution for Operational Items

Flag	Reading/Language Arts						Mathematics					
	Grade 2		Grade 3		Grade 4		Grade 2		Grade 3		Grade 4	
	N	Pct.	N	Pct.	N	Pct.	N	Pct.	N	Pct.	N	Pct.
A	14	22%	17	26%	25	33%	28	43%	26	40%	22	34%
B	14	22%	20	31%	19	25%	22	34%	15	23%	20	31%
C	23	35%	24	37%	23	31%	12	18%	22	34%	23	35%
D	10	15%	4	6%	4	5%	2	3%	1	2%	0	0%
F	4	6%	0	0%	4	5%	1	2%	1	2%	0	0%
TOTAL	65	100%	65	100%	75	100%	65	100%	65	100%	65	100%

Table 7.B.10 IRT Model Data Fit Distribution for Field-test Items

Flag	Reading/Language Arts						Mathematics					
	Grade 2		Grade 3		Grade 4		Grade 2		Grade 3		Grade 4	
	N	Pct.	N	Pct.	N	Pct.	N	Pct.	N	Pct.	N	Pct.
A	12	17%	16	22%	21	29%	22	31%	8	11%	25	35%
B	9	13%	11	15%	16	22%	25	36%	14	19%	11	15%
C	23	32%	25	35%	23	32%	20	29%	22	31%	20	28%
D	16	23%	8	11%	5	7%	1	1%	12	17%	7	10%
F	11	15%	12	17%	7	10%	2	3%	16	22%	9	13%
TOTAL	71	100%	72	100%	72	100%	70	100%	72	100%	72	100%

Chapter 8: Reliability

This chapter summarizes the evidence of reliability for the California Standards-based Tests in Spanish for the spring 2007 administration. These analyses were conducted for each of the six operational STS tests. The reliability analyses included the computation of overall and subscore reliabilities, standard errors of measurement (SEMs), and inter-correlations of reporting cluster subscores for the target population. Reliability analyses were reported both for the target population and at the subgroup level within the target population.

Reliability

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 distributions 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 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 reported below:

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_t^2} \right] \quad (8.1)$$

where,

n is the number of items,

σ_i^2 is the variance of scores on the i -th item, and

σ_t^2 is the variance of the total score (either the total raw score or scale score).

The SEM provides a measure of score instability in the score metric. Students' true abilities will fall within a range of their observed score ± 1 SEM about 68 percent of the time. The SEM was computed as follows:

$$\sigma_e = \sigma_t \sqrt{1 - \alpha} \quad (8.2)$$

where,

α is the reliability estimated using (8.1) above, and

σ_t^2 is the standard deviation of the total raw scores.

The reliability analyses were conducted for all valid cases of the target examinee population. Table 8.1 presents the results of reliability analyses on each of the six operational STS tests along with the number of items and examinees upon which those analyses were performed. The results in Table 8.1 indicated that all STS tests were highly reliable, with reliabilities ranging from 0.90 to 0.93. The reliabilities were highly consistent across all STS tests within the subject areas of RLA and mathematics.

Table 8.1 Reliabilities and Standard Errors of Measurement for the STS

Subject	STS	No. of Items	No. of Examinees	Reliability	Raw Score		
					Mean	Std. Dev.	SEM
Reading/Language Arts	2	65	12,486	0.92	40.82	12.05	3.41
	3	65	8,952	0.90	37.07	11.34	3.59
	4	75	5,291	0.92	40.47	13.55	3.83
Mathematics	2	65	12,469	0.92	44.84	11.33	3.20
	3	65	8,926	0.93	43.78	12.46	3.30
	4	65	5,283	0.93	39.93	12.90	3.41

Inter-correlations, Reliabilities, and SEMs for Reporting Clusters

For each STS test, number-correct scores are computed for five reporting clusters. Inter-correlations and reliability estimates for the reporting clusters are presented in Table 8.A.1 and Table 8.A.2 for the six STS tests. As expected, the reliabilities across reporting clusters varied significantly according to the number of items in each cluster. For example, the reliabilities for the first four reporting clusters in Math grade 4 were moderately high, ranging from 0.68 to 0.85. However, the fifth reporting cluster (Statistics, Data Analysis, and Probability) consisting of only four items had a coefficient alpha of 0.38. Similar results were observed for inter-correlations among reporting clusters, that is, the reporting clusters based on fewer items tended to have lower inter-correlations with the other clusters.

Subgroup Reliabilities and SEMs

The reliabilities of the six operational STS tests were also examined for various subgroups of the examinee population. The subgroups included in these analyses were gender, enrollment in the NSLP, provision of special services, length of attendance in U.S. schools, and EL program participation. For subgroups with fewer than 11 examinees, reliability and SEM results were not presented.

For each subgroup analysis, reliability and SEM information is first reported for the test overall and then at the cluster score level. The tables also include the corresponding sample sizes and the number of items used to compute those reliabilities and errors of measurement. Table 8.A.3 through Table 8.A.7 present the overall reliabilities, while Table 8.A.8 through Table 8.A.13 present the cluster-based reliabilities for the various subgroups. Table 8.A.8 and Table 8.A.9 present the cluster-based reliabilities for the subgroups based on Gender and NSLP enrollment. Table 8.A.10 and Table 8.A.11 show the same analyses for the subgroups based on provision of special services and attendance in U.S. schools. The last two tables, Table 8.A.12 through Table 8.A.13, present results for the subgroups based on EL program participation.

References

Cronbach, L.J. "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika*, Vol. 16, pp. 292–334, 1951.

Feldt, L. S. and Brennan, R. L. "Reliability," in *Educational Measurement*. Edited by R. L. Linn. New York: Macmillan, 1989.

Livingston, S. A., and Lewis, C. "Estimating the Consistency and Accuracy of Classification Based on Test Scores," *Journal on Educational Measurement*, Vol. 32, pp. 179–197, 1995.

Appendix 8.A—Reliabilities and SEM Tables

Table 8.A.1 Subscore Reliabilities and Correlations for Reading/Language Arts

Subscore Area	No. of Items	Correlation					Reliab.	SEM
		1	2	3	4	5		
Grade 2								
1. Word Analysis and Vocabulary Development	22	1.00	0.74	0.66	0.74	0.49	0.81	1.87
2. Reading Comprehension	15	0.74	1.00	0.63	0.69	0.49	0.77	1.64
3. Literary Response and Analysis	6	0.66	0.63	1.00	0.61	0.41	0.59	0.94
4. Written Conventions	14	0.74	0.69	0.61	1.00	0.52	0.79	1.55
5. Writing Strategies	8	0.49	0.49	0.41	0.52	1.00	0.39	1.33
Grade 3								
1. Word Analysis and Vocabulary Development	20	1.00	0.66	0.57	0.65	0.61	0.78	1.80
2. Reading Comprehension	15	0.66	1.00	0.57	0.62	0.60	0.67	1.75
3. Literary Response and Analysis	8	0.57	0.57	1.00	0.52	0.49	0.47	1.21
4. Written Conventions	13	0.65	0.62	0.52	1.00	0.62	0.68	1.65
5. Writing Strategies	9	0.61	0.60	0.49	0.62	1.00	0.60	1.36
Grade 4								
1. Word Analysis and Vocabulary Development	18	1.00	0.69	0.53	0.73	0.67	0.77	1.85
2. Reading Comprehension	15	0.69	1.00	0.54	0.64	0.61	0.71	1.74
3. Literary Response and Analysis	9	0.53	0.54	1.00	0.51	0.48	0.44	1.36
4. Written Conventions	18	0.73	0.64	0.51	1.00	0.73	0.79	1.84
5. Writing Strategies	15	0.67	0.61	0.48	0.73	1.00	0.67	1.74

Table 8.A.2 Subscore Reliabilities and Correlations for Mathematics

Subscore Area	No. of Items	Correlation					Reliab.	SEM
		1	2	3	4	5		
Grade 2								
1. Place Value, Addition, and Subtraction	15	1.00	0.74	0.59	0.59	0.63	0.76	1.58
2. Multiplication, Division, and Fractions	23	0.74	1.00	0.60	0.63	0.64	0.81	1.94
3. Algebra and Functions	6	0.59	0.60	1.00	0.47	0.51	0.61	1.03
4. Measurement and Geometry	14	0.59	0.63	0.47	1.00	0.56	0.67	1.45
5. Statistics, Data Analysis, and Probability	7	0.63	0.64	0.51	0.56	1.00	0.63	1.01
Grade 3								
1. Place Value, Fractions, and Decimals	16	1.00	0.78	0.73	0.67	0.58	0.78	1.64
2. Addition, Subtraction, Multiplication, Division	16	0.78	1.00	0.74	0.65	0.56	0.81	1.60
3. Algebra and Functions	12	0.73	0.74	1.00	0.63	0.56	0.76	1.43
4. Measurement and Geometry	16	0.67	0.65	0.63	1.00	0.58	0.71	1.62
5. Statistics, Data Analysis, and Probability	5	0.58	0.56	0.56	0.58	1.00	0.66	0.80
Grade 4								
1. Decimals, Fractions, and Negative Numbers	17	1.00	0.69	0.69	0.55	0.47	0.77	1.73
2. Operations and Factoring	14	0.69	1.00	0.71	0.55	0.47	0.78	1.60
3. Algebra and Functions	18	0.69	0.71	1.00	0.61	0.50	0.85	1.76
4. Measurement and Geometry	12	0.55	0.55	0.61	1.00	0.43	0.68	1.51
5. Statistics, Data Analysis, and Probability	4	0.47	0.47	0.50	0.43	1.00	0.38	0.90

Table 8.A.3 Reliabilities and SEM for the STS by GENDER

Subject	STS	Male			Female		
		N	Reliab.	SEM	N	Reliab.	SEM
Reading/Language Arts	2	6,323	0.92	3.44	6,163	0.92	3.32
	3	4,481	0.90	3.57	4,471	0.90	3.52
	4	2,615	0.92	3.81	2,676	0.92	3.71
Mathematics	2	6,316	0.92	3.24	6,153	0.92	3.17
	3	4,472	0.93	3.40	4,454	0.93	3.19
	4	2,607	0.93	3.50	2,676	0.92	3.54

Table 8.A.4 Reliabilities and SEM for the STS by NSLP

Subject	STS	Not in NSLP			NSLP		
		N	Reliab.	SEM	N	Reliab.	SEM
Reading/Language Arts	2	1,133	0.93	3.40	11,324	0.92	3.39
	3	942	0.91	3.55	7,989	0.90	3.57
	4	565	0.92	3.86	4,704	0.92	3.83
Mathematics	2	1,133	0.93	3.18	11,307	0.92	3.19
	3	941	0.93	3.35	7,965	0.93	3.29
	4	565	0.93	3.43	4,696	0.93	3.41

Table 8.A.5 Reliabilities and SEM for the STS by SPECIAL EDUCATION

Subject	STS	No Special Education			Special Education		
		N	Reliab.	SEM	N	Reliab.	SEM
Reading/Language Arts	2	11,879	0.92	3.38	607	0.91	3.53
	3	8,484	0.90	3.55	468	0.88	3.68
	4	5,073	0.92	3.79	218	0.91	3.83
Mathematics	2	11,854	0.92	3.17	615	0.92	3.48
	3	8,464	0.93	3.26	462	0.93	3.41
	4	5,061	0.93	3.39	222	0.93	3.50

Table 8.A.6 Reliabilities and SEM for the STS by ATTENDANCE IN U.S. SCHOOLS

Subject	STS	In U.S. Schools < 12 Months			In U.S. Schools >= 12 Months		
		N	Reliab.	SEM	N	Reliab.	SEM
Reading/Language Arts	2	1,898	0.93	3.48	10,588	0.91	3.39
	3	1,709	0.90	3.60	7,243	0.90	3.52
	4	1,573	0.91	3.91	3,718	0.92	3.89
Mathematics	2	1,906	0.93	3.33	10,563	0.91	3.23
	3	1,711	0.93	3.48	7,215	0.93	3.14
	4	1,572	0.92	3.61	3,711	0.93	3.34

Table 8.A.7 Reliabilities and SEM for the STS by EL PROGRAM PARTICIPATION

Subject	STS	EL in ELD			EL in ELD and SDAIE			EL in ELD and SDAIE with Primary Language Support			EL in ELD and Academic Subjects through Primary Language			Other EL Instructional 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	165	0.91	3.68	576	0.91	3.70	629	0.92	3.57	11,008	0.91	3.40	12	0.96	3.13	73	0.92	3.54
	3	154	0.90	3.66	582	0.90	3.64	577	0.89	3.61	7,555	0.90	3.51	11	0.95	3.36	63	0.87	3.62
	4	163	0.91	3.96	590	0.92	3.82	578	0.91	3.82	3,896	0.92	3.87	5	N/A	N/A	43	0.90	3.81
Mathematics	2	165	0.91	3.58	583	0.92	3.46	632	0.92	3.52	10,982	0.91	3.25	13	0.94	3.40	72	0.92	3.48
	3	154	0.93	3.57	587	0.93	3.49	573	0.93	3.43	7,526	0.93	3.15	12	0.95	3.20	64	0.94	3.36
	4	163	0.92	3.58	591	0.93	3.52	577	0.91	3.67	3,888	0.93	3.33	5	N/A	N/A	43	0.92	3.53

Table 8.A.8 Subscore Reliabilities and SEM for Reading/Language Arts by GENDER/NSLP

Subscore Area	No. of Items	Male			Female			Not in NSLP		
		Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	
Grade 2										
1. Word Analysis and Vocabulary Development	22	0.80	1.91	0.80	1.83	0.83	1.88	0.80	1.88	1.87
2. Reading Comprehension	15	0.76	1.68	0.77	1.59	0.79	1.63	0.76	1.63	1.64
3. Literary Response and Analysis	6	0.59	0.95	0.58	0.91	0.64	0.93	0.59	0.93	0.94
4. Written Conventions	14	0.78	1.57	0.78	1.52	0.81	1.53	0.78	1.53	1.55
5. Writing Strategies	8	0.39	1.32	0.39	1.33	0.40	1.33	0.39	1.33	1.33
Grade 3										
1. Word Analysis and Vocabulary Development	20	0.78	1.84	0.77	1.77	0.79	1.80	0.78	1.80	1.80
2. Reading Comprehension	15	0.66	1.76	0.67	1.74	0.70	1.74	0.67	1.74	1.75
3. Literary Response and Analysis	8	0.46	1.22	0.48	1.20	0.56	1.18	0.46	1.18	1.21
4. Written Conventions	13	0.68	1.66	0.68	1.64	0.71	1.63	0.68	1.63	1.65
5. Writing Strategies	9	0.58	1.37	0.60	1.35	0.62	1.35	0.59	1.35	1.36
Grade 4										
1. Word Analysis and Vocabulary Development	18	0.77	1.88	0.77	1.82	0.78	1.84	0.77	1.84	1.85
2. Reading Comprehension	15	0.70	1.75	0.71	1.72	0.72	1.73	0.71	1.73	1.74
3. Literary Response and Analysis	9	0.43	1.36	0.42	1.37	0.42	1.38	0.44	1.38	1.36
4. Written Conventions	18	0.78	1.87	0.78	1.81	0.78	1.84	0.79	1.84	1.84
5. Writing Strategies	15	0.68	1.75	0.66	1.73	0.68	1.73	0.67	1.73	1.74

Table 8.A.9 Subscore Reliabilities and SEM for Mathematics by GENDER/NSLP

Subscore Area	No. of Items	Male		Female		Not in NSLP		NSLP	
		Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
Grade 2									
1. Place Value, Addition, and Subtraction	15	0.76	1.57	0.76	1.58	0.78	1.57	0.76	1.58
2. Multiplication, Division, and Fractions	23	0.81	1.95	0.80	1.93	0.82	1.93	0.80	1.94
3. Algebra and Functions	6	0.61	1.02	0.60	1.03	0.63	1.02	0.60	1.03
4. Measurement and Geometry	14	0.68	1.44	0.67	1.47	0.71	1.43	0.67	1.46
5. Statistics, Data Analysis, and Probability	7	0.63	1.02	0.63	1.01	0.65	1.03	0.62	1.01
Grade 3									
1. Place Value, Fractions, and Decimals	16	0.79	1.63	0.77	1.64	0.78	1.64	0.78	1.64
2. Addition, Subtraction, Multiplication, Division	16	0.81	1.61	0.80	1.58	0.81	1.58	0.81	1.60
3. Algebra and Functions	12	0.78	1.41	0.74	1.44	0.77	1.42	0.76	1.43
4. Measurement and Geometry	16	0.74	1.62	0.69	1.62	0.74	1.61	0.71	1.62
5. Statistics, Data Analysis, and Probability	5	0.66	0.82	0.66	0.78	0.67	0.80	0.66	0.80
Grade 4									
1. Decimals, Fractions, and Negative Numbers	17	0.78	1.73	0.75	1.74	0.78	1.72	0.77	1.74
2. Operations and Factoring	14	0.78	1.62	0.79	1.58	0.79	1.59	0.78	1.60
3. Algebra and Functions	18	0.86	1.76	0.84	1.75	0.85	1.77	0.85	1.76
4. Measurement and Geometry	12	0.69	1.52	0.67	1.50	0.70	1.50	0.68	1.51
5. Statistics, Data Analysis, and Probability	4	0.41	0.90	0.35	0.89	0.41	0.89	0.38	0.90

Table 8.A.10 Subscore Reliabilities and SEM for Reading/Language Arts by SPECIAL EDUCATION/ATTENDANCE IN U.S. SCHOOLS

Subscore Area	No. of Items	No Special Education		Special Education		In U.S. Schools < 12 Months		In U.S. Schools ≥ 12 Months	
		Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
Grade 2									
1. Word Analysis and Vocabulary Development	22	0.80	1.87	0.78	2.04	0.82	2.02	0.78	1.84
2. Reading Comprehension	15	0.77	1.63	0.72	1.75	0.80	1.68	0.75	1.63
3. Literary Response and Analysis	6	0.59	0.93	0.58	1.01	0.64	1.03	0.55	0.92
4. Written Conventions	14	0.79	1.54	0.73	1.67	0.77	1.65	0.77	1.53
5. Writing Strategies	8	0.39	1.33	0.34	1.31	0.36	1.30	0.39	1.33
Grade 3									
1. Word Analysis and Vocabulary Development	20	0.77	1.80	0.78	1.94	0.79	1.89	0.77	1.78
2. Reading Comprehension	15	0.67	1.75	0.60	1.76	0.66	1.77	0.67	1.74
3. Literary Response and Analysis	8	0.47	1.21	0.39	1.24	0.53	1.20	0.45	1.21
4. Written Conventions	13	0.68	1.65	0.62	1.67	0.68	1.64	0.67	1.65
5. Writing Strategies	9	0.59	1.36	0.50	1.36	0.55	1.36	0.59	1.36
Grade 4									
1. Word Analysis and Vocabulary Development	18	0.77	1.85	0.74	1.93	0.75	1.84	0.79	1.84
2. Reading Comprehension	15	0.71	1.74	0.68	1.73	0.68	1.72	0.73	1.74
3. Literary Response and Analysis	9	0.43	1.37	0.43	1.32	0.35	1.40	0.47	1.35
4. Written Conventions	18	0.78	1.84	0.74	1.90	0.79	1.86	0.78	1.83
5. Writing Strategies	15	0.67	1.74	0.68	1.73	0.67	1.76	0.68	1.73

Table 8.A.11 Subscore Reliabilities and SEM for Mathematics by SPECIAL EDUCATION/ATTENDANCE IN U.S. SCHOOLS

Subscore Area	No. of Items	No Special Education		Special Education		In U.S. School < 12 Months		In U.S. Schools >= 12 Months	
		Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
Grade 2									
1. Place Value, Addition, and Subtraction	15	0.76	1.57	0.77	1.65	0.77	1.66	0.74	1.56
2. Multiplication, Division, and Fractions	23	0.80	1.93	0.81	2.02	0.82	2.00	0.80	1.92
3. Algebra and Functions	6	0.60	1.02	0.58	1.08	0.59	1.07	0.60	1.02
4. Measurement and Geometry	14	0.67	1.45	0.70	1.58	0.73	1.55	0.65	1.43
5. Statistics, Data Analysis, and Probability	7	0.62	1.01	0.66	1.09	0.63	1.12	0.61	0.99
Grade 3									
1. Place Value, Fractions, and Decimals	16	0.78	1.63	0.78	1.71	0.78	1.72	0.77	1.61
2. Addition, Subtraction, Multiplication, Division	16	0.81	1.59	0.80	1.70	0.79	1.68	0.81	1.57
3. Algebra and Functions	12	0.76	1.42	0.74	1.51	0.75	1.49	0.75	1.41
4. Measurement and Geometry	16	0.71	1.61	0.72	1.73	0.74	1.71	0.68	1.59
5. Statistics, Data Analysis, and Probability	5	0.65	0.79	0.67	0.87	0.63	0.93	0.63	0.76
Grade 4									
1. Decimals, Fractions, and Negative Numbers	17	0.77	1.73	0.77	1.82	0.77	1.78	0.76	1.71
2. Operations and Factoring	14	0.78	1.60	0.76	1.64	0.78	1.63	0.78	1.59
3. Algebra and Functions	18	0.85	1.76	0.86	1.81	0.83	1.85	0.85	1.72
4. Measurement and Geometry	12	0.68	1.51	0.67	1.56	0.65	1.55	0.69	1.49
5. Statistics, Data Analysis, and Probability	4	0.38	0.90	0.38	0.90	0.38	0.91	0.37	0.89

Table 8.A.12 Subscore Reliabilities and SEM for Reading/Language Arts by EL PROGRAM PARTICIPATION

Subscore Area	No. of Items	EL in ELD		EL in ELD and SDAIE		EL in ELD and SDAIE with Primary Lang. Support		EL in ELD and Academic Subj. through Primary Lang.		Other EL Instructional Services		None (EL only)	
		Rel.	SEM	Rel.	SEM	Rel.	SEM	Rel.	SEM	Rel.	SEM	Rel.	SEM
Grade 2													
1. Word Analysis and Vocabulary Development	22	0.79	2.06	0.80	2.07	0.80	2.07	0.79	1.84	0.92	1.72	0.82	2.00
2. Reading Comprehension	15	0.80	1.68	0.78	1.70	0.78	1.69	0.75	1.63	0.82	1.59	0.81	1.64
3. Literary Response and Analysis	6	0.66	1.01	0.64	1.04	0.60	1.07	0.55	0.92	0.52	1.04	0.52	1.07
4. Written Conventions	14	0.70	1.69	0.72	1.67	0.74	1.66	0.77	1.53	0.86	1.49	0.80	1.61
5. Writing Strategies	8	0.40	1.27	0.27	1.31	0.30	1.28	0.39	1.33	0.67	1.20	0.35	1.30
Grade 3													
1. Word Analysis and Vocabulary Development	20	0.79	1.92	0.79	1.90	0.78	1.92	0.77	1.78	0.89	1.69	0.74	1.87
2. Reading Comprehension	15	0.71	1.74	0.65	1.77	0.61	1.77	0.67	1.74	0.80	1.64	0.56	1.80
3. Literary Response and Analysis	8	0.57	1.20	0.52	1.19	0.53	1.21	0.46	1.21	0.72	1.11	0.54	1.21
4. Written Conventions	13	0.70	1.61	0.69	1.64	0.64	1.65	0.67	1.65	0.87	1.39	0.64	1.63
5. Writing Strategies	9	0.48	1.38	0.62	1.34	0.49	1.37	0.59	1.36	0.84	1.11	0.46	1.37
Grade 4													
1. Word Analysis and Vocabulary Development	18	0.77	1.85	0.77	1.83	0.73	1.85	0.78	1.84	N/A	N/A	0.72	1.86
2. Reading Comprehension	15	0.71	1.70	0.69	1.73	0.66	1.72	0.72	1.74	N/A	N/A	0.67	1.73
3. Literary Response and Analysis	9	0.44	1.39	0.35	1.40	0.30	1.40	0.47	1.35	N/A	N/A	0.29	1.37
4. Written Conventions	18	0.77	1.88	0.80	1.86	0.79	1.87	0.78	1.83	N/A	N/A	0.77	1.82
5. Writing Strategies	15	0.69	1.73	0.67	1.76	0.66	1.76	0.68	1.73	N/A	N/A	0.62	1.74

Table 8.A.13 Subscore Reliabilities and SEM for Mathematics by EL PROGRAM PARTICIPATION

Subscore Area	No. of Items	EL in ELD		EL in ELD and SDAIE		EL in ELD and SDAIE with Primary Lang. Support		EL in ELD and Academic Subj. through Primary Lang.		Other EL Instructional Services		None (EL only)	
		Rel.	SEM	Rel.	SEM	Rel.	SEM	Rel.	SEM	Rel.	SEM	Rel.	SEM
Grade 2													
1. Place Value, Addition, and Subtraction	15	0.73	1.69	0.74	1.68	0.75	1.69	0.75	1.56	0.68	1.65	0.80	1.63
2. Multiplication, Division, and Fractions	23	0.80	2.03	0.81	2.02	0.81	2.02	0.80	1.92	0.88	1.91	0.79	2.02
3. Algebra and Functions	6	0.58	1.07	0.59	1.08	0.57	1.08	0.60	1.02	0.66	1.01	0.55	1.09
4. Measurement and Geometry	14	0.67	1.61	0.74	1.56	0.73	1.58	0.65	1.43	0.81	1.52	0.71	1.53
5. Statistics, Data Analysis, and Probability	7	0.62	1.13	0.62	1.13	0.60	1.13	0.61	0.99	0.63	1.12	0.58	1.12
Grade 3													
1. Place Value, Fractions, and Decimals	16	0.81	1.70	0.78	1.71	0.77	1.73	0.77	1.62	0.76	1.72	0.82	1.70
2. Addition, Subtraction, Multiplication, Division	16	0.80	1.69	0.79	1.67	0.79	1.71	0.81	1.58	0.88	1.53	0.83	1.66
3. Algebra and Functions	12	0.71	1.51	0.76	1.48	0.75	1.49	0.75	1.42	0.86	1.31	0.73	1.50
4. Measurement and Geometry	16	0.76	1.71	0.74	1.71	0.72	1.74	0.69	1.60	0.70	1.75	0.77	1.70
5. Statistics, Data Analysis, and Probability	5	0.66	0.92	0.62	0.92	0.61	0.95	0.64	0.77	0.78	0.77	0.49	0.97
Grade 4													
1. Decimals, Fractions, and Negative Numbers	17	0.78	1.76	0.80	1.76	0.72	1.81	0.76	1.71	N/A	N/A	0.78	1.77
2. Operations and Factoring	14	0.80	1.59	0.79	1.62	0.75	1.65	0.78	1.59	N/A	N/A	0.80	1.60
3. Algebra and Functions	18	0.82	1.86	0.84	1.85	0.83	1.86	0.85	1.72	N/A	N/A	0.83	1.83
4. Measurement and Geometry	12	0.61	1.56	0.65	1.55	0.66	1.54	0.68	1.49	N/A	N/A	0.63	1.53
5. Statistics, Data Analysis, and Probability	4	0.41	0.89	0.44	0.89	0.33	0.92	0.36	0.90	N/A	N/A	0.29	0.92