California Department of Education Assessment Development and Administration Division

California Alternate Performance Assessment

Technical Report

Spring 2011 Administration

Draft Submitted March 22, 2012

Educational Testing Service

Contract No. 5417

Table of Contents

Acronyms and Initialisms Used in the *CAPA Technical Report* vi

Chapter 1: Introduction 1
Background 1
Test Purpose 1
Test Content 1
Intended Population 2
Intended Use and Purpose of Test Scores 3
Testing Window 3
Significant STAR Developments in 2011 4
Limitations of the Assessment 4
Score Interpretation 4
Out-of-Level Testing 4
Score Comparison 4
Groups and Organizations Groups Involved in Test Development 4
State Board of Education 4
California Department of Education 5
Contractors 5
Overview of the Technical Report 5
Reference 7
Chapter 2: An Overview of CAPA Processes 8
Task (Item) Development 8
Task Formats 8
Task (Item) Specifications 8
Item Banking 8
Task Refresh Rate 9
Test Assembly 9
Test Length 9
Test Blueprints 9
Content Rules and Task Selection 9
Psychometric Criteria 9
Test Administration 10
Test Security and Confidentiality 10
Procedures to Maintain Standardization 10
Test Variations, Accommodations, and Modifications 11
Scores 11
Aggregation Procedures 12
Equating 12
Calibration 12
Scaling 13
Linear Transformation 13
References 15
Chapter 3: Task (Item) Development 16
Rules for Task Development 16
Task Specifications 16
Expected Task Ratio 17
Selection of Task Writers 18
Criteria for Selecting Task Writers 18
Task (Item) Review Process 18
Contractor Review 18
Content Expert Reviews 20
Statewide Pupil Assessment Review Panel 22
Field Testing 22
Stand-alone Field Testing 22
Embedded Field-test Tasks 22
CDE Data Review 24
Item Banking 24
References 25

Chapter 4: Test Assembly 26
Test Length 26
Rules for Task Selection 26
Test Blueprints 26
Content Rules and Task Selection 26
Psychometric Criteria 27
Projected Psychometric Properties of the Assembled Tests 28
Rules for Task Sequence and Layout 28
Chapter 5: Test Administration 29
Test Security and Confidentiality 29
ETS's Office of Testing Integrity 29
Test Development 29
Task and Data Review 29
Item Banking 30
Transfer of Forms and Tasks to the CDE 30
Security of Electronic Files Using a Firewall 30
Printing and Publishing 31
Test Administration 31
Test Delivery 31
Processing and Scoring 32
Data Management 32
Transfer of Scores via Secure Data Exchange 33
Statistical Analysis 33
Reporting and Posting Results 33
Student Confidentiality 33
Student Test Results 33
Procedures to Maintain Standardization 34
Test Administrators 34
CAPA Examiner's Manual 35
District and Test Site Coordinator Manual 35
STAR Management System Manuals 36
Accommodations for Students with Disabilities 36
Identification 36
Adaptations 36
Scoring 37
Demographic Data Corrections 37
Testing Irregularities 37
Test Administration Incidents 37
References 39
Chapter 6: Performance Standards 40
Background 40
Standard Setting Procedure 40
Development of Competencies Lists 41
Standard Setting Methodology 42
Performance Profile Method 42
Results 42
References 44
Chapter 7: Scoring and Reporting 45
Procedures for Maintaining and Retrieving Individual Scores 45
Scoring and Reporting Specifications 46
Scanning and Scoring 46
Types of Scores 47
Raw Score 47
Scale Score 47
Performance Levels 47
Score Verification Procedures 47
Monitoring and Quality Control of Scoring 47
Score Verification Process 48
Overview of Score Aggregation Procedures 48
Individual Scores 48
Score Distributions and Summary Statistics 48
Reports Produced and Scores for Each Report 53
Types of Score Reports 53

Score Report Contents 53
Score Report Applications 53
Criteria for Interpreting Test Scores 54
Criteria for Interpreting Score Reports 54
References 56
Appendix 7.A—Scale Score Distribution Tables 57
Appendix 7.B—Demographic Summaries 59
Appendix 7.C—Type of Score Report 65
Chapter 8: Analyses 68
Samples Used for the Analyses 68
Classical Analyses 69
Average Item Score 69
Polyserial Correlation of the Task Score with the Total Test Score 69
Reliability Analyses 71
Subgroup Reliabilities and SEMs 72
Conditional Standard Errors of Measurement 72
Decision Classification Analyses 73
Validity Evidence 74
Purpose of the CAPA 75
The Constructs to Be Measured 75
Interpretations and Uses of the Scores Generated 75
Intended Test Population(s) 76
Validity Evidence Collected 76
Evidence Based on Response Processes 79
Evidence of Interrater Agreement 79
Evidence Based on Internal Structure 80
Evidence Based on Consequences of Testing 80
IRT Analyses 80
IRT Model-Data Fit Analyses 81
Model-fit Assessment Results 82
Evaluation of Scaling 82
Summaries of Scaled IRT b-values 83
Post-scaling Results 83
Differential Item Functioning Analyses 83
References 86
Appendix 8.A—Classical Analyses: Task Statistics 88
Appendix 8.B—Reliability Analyses 105
Appendix 8.C—Correlations Between Content Area Scores 119
Appendix 8.D—IRT Analyses 138
Appendix 8.E—DIF Analyses 154
Chapter 9: Quality Control Procedures 160
Quality Control of Task Development 160
Task Specifications 160
Task Writers 160
Internal Contractor Reviews 160
Assessment Review Panel Review 161
Statewide Pupil Assessment Review Panel Review 161
Data Review of Field-tested Tasks 161
Quality Control of the Item Bank 162
Quality Control of Test Form Development 162
Quality Control of Test Materials 163
Collecting Test Materials 163
Processing Test Materials 163
Quality Control of Scanning 163
Post-scanning Edits 164
Quality Control of Image Editing 164
Quality Control of Answer Document Processing and Scoring 164
Accountability of Answer Documents 164
Processing of Answer Documents 165
Scoring and Reporting Specifications 165
Matching Information on CAPA Answer Documents 165
Storing Answer Documents 165

Quality Control of Psychometric Processes **166**
Quality Control of Task (Item) Analyses, DIF, and the Scoring Process 166
Score Verification Process 167
Year-to-Year Comparison Analyses 167
Offloads to Test Development 167
Quality Control of Reporting **167**
Excluding Student Scores from Summary Reports 168
Reference **169**
Chapter 10: Historical Comparisons **170**
Base Year Comparisons **170**
Examinee Performance **170**
Test Characteristics **171**
Appendix 10.A—Historical Comparisons Tables **172**
Appendix 10.B—Historical Comparisons Tables **176**

Tables

Table 1.1 Description of the CAPA Assessment Levels 2
Table 2.1 CAPA Item and Estimated Time Chart 9
Table 2.2 Scale Scores Ranges for Performance Levels 14
Table 3.1 Field-test Percentages for the CAPA 17
Table 3.2 CAPA ARP Member Qualifications, by Content Area and Total 21
Table 3.3 Summary of Tasks and Forms Presented in the 2011 CAPA 23
Table 4.1 Target Statistical Specifications for the CAPA 27
Table 4.2 Summary of 2011 CAPA Projected Statistical Attributes 28
Table 7.1 Rubrics for CAPA Scoring 45
Table 7.2 Summary Statistics Describing Student Scores: ELA 48
Table 7.3 Summary Statistics Describing Student Scores: Mathematics 49
Table 7.4 Summary Statistics Describing Student Scores: Science 50
Table 7.5 Percentage of Examinees in Each Performance Level 51
Table 7.6 Subgroup Definitions 52
Table 7.7 Types of CAPA Reports 53
Table 7.A.1 Scale Score Frequency Distributions: ELA, Levels I–V 57
Table 7.A.2 Scale Score Frequency Distributions: Mathematics, Levels I–V 57
Table 7.A.3 Scale Score Frequency Distributions: Science, Levels I–V 58
Table 7.B.1 Demographic Summary for ELA, All Examinees 59
Table 7.B.2 Demographic Summary for Mathematics, All Examinees 61
Table 7.B.3 Demographic Summary for Science, All Examinees 63
Table 7.C.1 Score Reports Reflecting CAPA Results 65
Table 8.1 CAPA Raw Score Means and Standard Deviations: Total P2 Population and Equating Sample 69
Table 8.2 Average Item Score and Polyserial Correlation 70
Table 8.3 Reliabilities and Standard Errors of Measurement for the CAPA 72
Table 8.4 CAPA Content-area Correlations for CAPA Levels 79
Table 8.5 Evaluation of Common Items Between New and Reference Test Forms 82
Table 8.6 DIF Flags Based on the ETS DIF Classification Scheme 85
Table 8.7 Subgroup Classification for DIF Analyses 85
Table 8.A.1 AIS and Polyserial Correlation: Level I, ELA 88
Table 8.A.2 AIS and Polyserial Correlation: Level II, ELA 89
Table 8.A.3 AIS and Polyserial Correlation: Level III, ELA 90
Table 8.A.4 AIS and Polyserial Correlation: Level IV, ELA 91
Table 8.A.5 AIS and Polyserial Correlation: Level V, ELA 92
Table 8.A.6 AIS and Polyserial Correlation: Level I, Mathematics 93
Table 8.A.7 AIS and Polyserial Correlation: Level II, Mathematics 94
Table 8.A.8 AIS and Polyserial Correlation: Level III, Mathematics 95
Table 8.A.9 AIS and Polyserial Correlation: Level IV, Mathematics 96
Table 8.A.10 AIS and Polyserial Correlation: Level V, Mathematics 97
Table 8.A.11 AIS and Polyserial Correlation: Level I, Science 98
Table 8.A.12 AIS and Polyserial Correlation: Level III, Science 99
Table 8.A.13 AIS and Polyserial Correlation: Level IV, Science 100
Table 8.A.14 AIS and Polyserial Correlation: Level V, Science 101
Table 8.A.15 Frequency of Operational Task Scores: ELA 102
Table 8.A.16 Frequency of Operational Task Scores: Mathematics 103
Table 8.A.17 Frequency of Operational Task Scores: Science 104
Table 8.B.1 Reliabilities and SEMs by GENDER 105
Table 8.B.2 Reliabilities and SEMs by PRIMARY ETHNICITY 106

Table 8.B.3 Reliabilities and SEMs by PRIMARY ETHNICITY for Economically Disadvantaged 107
Table 8.B.4 Reliabilities and SEMs by PRIMARY ETHNICITY for Not Economically Disadvantaged..... 108
Table 8.B.5 Reliabilities and SEMs by PRIMARY ETHNICITY for Unknown Economic Status 109
Table 8.B.6 Reliabilities and SEMs by Disability 110
Table 8.B.7 Decision Accuracy and Decision Consistency: Level I, ELA..... 111
Table 8.B.8 Decision Accuracy and Decision Consistency: Level I, Mathematics 112
Table 8.B.9 Decision Accuracy and Decision Consistency: Level I, Science..... 112
Table 8.B.10 Decision Accuracy and Decision Consistency: Level II, ELA..... 113
Table 8.B.11 Decision Accuracy and Decision Consistency: Level II, Mathematics 113
Table 8.B.12 Decision Accuracy and Decision Consistency: Level III, ELA..... 114
Table 8.B.13 Decision Accuracy and Decision Consistency: Level III, Mathematics 114
Table 8.B.14 Decision Accuracy and Decision Consistency: Level III, Science..... 115
Table 8.B.15 Decision Accuracy and Decision Consistency: Level IV, ELA 115
Table 8.B.16 Decision Accuracy and Decision Consistency: Level IV, Mathematics..... 116
Table 8.B.17 Decision Accuracy and Decision Consistency: Level IV, Science 116
Table 8.B.18 Decision Accuracy and Decision Consistency: Level V, ELA 117
Table 8.B.19 Decision Accuracy and Decision Consistency: Level V, Mathematics..... 117
Table 8.B.20 Decision Accuracy and Decision Consistency: Level V, Science 118
Table 8.C.1 CAPA Content Area Correlations by Gender: Level I 119
Table 8.C.2 CAPA Content Area Correlations by Gender: Level II 119
Table 8.C.3 CAPA Content Area Correlations by Gender: Level III 119
Table 8.C.4 CAPA Content Area Correlations by Gender: Level IV..... 119
Table 8.C.5 CAPA Content Area Correlations by Gender: Level V..... 119
Table 8.C.6 CAPA Content Area Correlations by Ethnicity: Level I..... 120
Table 8.C.7 CAPA Content Area Correlations by Ethnicity: Level II..... 120
Table 8.C.8 CAPA Content Area Correlations by Ethnicity: Level III..... 120
Table 8.C.9 CAPA Content Area Correlations by Ethnicity: Level IV 121
Table 8.C.10 CAPA Content Area Correlations by Ethnicity: Level V 121
Table 8.C.11 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level I..... 121
Table 8.C.12 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level II..... 122
Table 8.C.13 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level III..... 122
Table 8.C.14 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level IV 122
Table 8.C.15 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level V 123
Table 8.C.16 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level I 123
Table 8.C.17 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level II 123
Table 8.C.18 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level III 124
Table 8.C.19 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level IV..... 124
Table 8.C.20 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level V..... 124
Table 8.C.21 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level I..... 125
Table 8.C.22 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level II..... 125
Table 8.C.23 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level III..... 125
Table 8.C.24 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level IV 126
Table 8.C.25 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level V 126
Table 8.C.26 CAPA Content Area Correlations by Economic Status: Level I 127
Table 8.C.27 CAPA Content Area Correlations by Economic Status: Level II 127
Table 8.C.28 CAPA Content Area Correlations by Economic Status: Level III 127
Table 8.C.29 CAPA Content Area Correlations by Economic Status: Level IV..... 127
Table 8.C.30 CAPA Content Area Correlations by Economic Status: Level V..... 127
Table 8.C.31 CAPA Content Area Correlations by Disability: Level I..... 128
Table 8.C.32 CAPA Content Area Correlations by Disability: Level II..... 129
Table 8.C.33 CAPA Content Area Correlations by Disability: Level III..... 130
Table 8.C.34 CAPA Content Area Correlations by Disability: Level IV 131
Table 8.C.35 CAPA Content Area Correlations by Disability: Level V 132
Table 8.C.36 Interrater Agreement Analyses for Operational Tasks: Level I 133
Table 8.C.37 Interrater Agreement Analyses for Operational Tasks: Level II 134
Table 8.C.38 Interrater Agreement Analyses for Operational Tasks: Level III 135
Table 8.C.39 Interrater Agreement Analyses for Operational Tasks: Level IV..... 136
Table 8.C.40 Interrater Agreement Analyses for Operational Tasks: Level V..... 137
Table 8.D.1 Item Classifications for Model-Data Fit Across All CAPA Levels 138
Table 8.D.2 Fit Classifications: Level I Tasks..... 138
Table 8.D.3 Fit Classifications: Level II Tasks..... 138
Table 8.D.4 Fit Classifications: Level III Tasks..... 138
Table 8.D.5 Fit Classifications: Level IV Tasks 138
Table 8.D.6 Fit Classifications: Level V Tasks 138
Table 8.D.7 IRT *b*-values for ELA, by Level..... 139

Table 8.D.8 IRT *b*-values for Mathematics, by Level 139
Table 8.D.9 IRT *b*-values for Science, by Level 139
Table 8.D.10 Score Conversions: Level I, ELA 140
Table 8.D.11 Score Conversions: Level II, ELA 141
Table 8.D.12 Score Conversions: Level III, ELA 142
Table 8.D.13 Score Conversions: Level IV, ELA 143
Table 8.D.14 Score Conversions: Level V, ELA 144
Table 8.D.15 Score Conversions: Level I, Mathematics 145
Table 8.D.16 Score Conversions: Level II, Mathematics 146
Table 8.D.17 Score Conversions: Level III, Mathematics 147
Table 8.D.18 Score Conversions: Level IV, Mathematics 148
Table 8.D.19 Score Conversions: Level V, Mathematics 149
Table 8.D.20 Score Conversions: Level I, Science 150
Table 8.D.21 Score Conversions: Level III, Science 151
Table 8.D.22 Score Conversions: Level IV, Science 152
Table 8.D.23 Score Conversions: Level V, Science 153
Table 8.E.1 Item Exhibiting Significant DIF by Ethnic Group 154
Table 8.E.2 Items Exhibiting Significant DIF by Disability Group 155
Table 8.E.3 CAPA Disability Distributions: Level I 157
Table 8.E.4 CAPA Disability Distributions: Level II 157
Table 8.E.5 CAPA Disability Distributions: Level III 158
Table 8.E.6 CAPA Disability Distributions: Level IV 158
Table 8.E.7 CAPA Disability Distributions: Level V 159
Table 10.A.1 Number of Examinees Tested, Scale Score Means and Standard Deviations of CAPA Across Base Year (2009), 2010, and 2011 172
Table 10.A.2 Percentage of Proficient and Above and Percentage of Advanced Across Base Year (2009), 2010, and 2011 172
Table 10.A.3 Observed Score Distributions of CAPA Across Base Year (2009), 2010, and 2011 for ELA 173
Table 10.A.4 Observed Score Distributions of CAPA Across Base Year (2009), 2010 and 2011 for Mathematics 174
Table 10.A.5 Observed Score Distributions of CAPA Across Base Year (2009), 2010, and 2011 for Science 175
Table 10.B.1 Average Item Score of CAPA Operational Test Items Across Base Year (2009), 2010, and 2011 176
Table 10.B.2 Mean IRT *b*-values for Operational Test Items Across Base Year (2009), 2010, and 2011 176
Table 10.B.3 Mean Polyserial Correlation of CAPA Operational Test Items Across Base Year (2009), 2010, and 2011 177
Table 10.B.4 Score Reliabilities and SEM of CAPA Across Base Year (2009) 2010, and 2011 177

Figures

Figure 3.1 The ETS Item Development Process for the STAR Program 16
Figure 8.1 Decision Accuracy for Achieving a Performance Level 74
Figure 8.2 Decision Consistency for Achieving a Performance Level 74

Acronyms and Initialisms Used in the *CAPA Technical Report*

1PPC	1-parameter partial credit
ADA	Americans with Disabilities Act
AIS	average task (item) score
API	Academic Performance Index
ARP	Assessment Review Panel
AYP	Adequate Yearly Progress
CAPA	California Alternate Performance Assessment
CCR	*California Code of Regulations*
CDE	California Department of Education
CDS	County-District-School
CI	confidence interval
CMA	California Modified Assessment
CSEMs	conditional standard errors of measurement
CSTs	California Standards Tests
DIF	Differential Task (Item) Functioning
DPLT	designated primary language test
DQS	Data Quality Services
EC	Education Code
EM	expectation maximization
ESEA	Elementary and Secondary Education Act
ETS	Educational Testing Service
GENASYS	Generalized Analysis System
HumRRo	Human Resource Research Organization
ICC	task (item) characteristic curve
IEP	individualized education program
IRF	item response functions
IRT	task (item) response theory
IT	Information Technology
LEA	local educational agency
MH	Mantel-Haenszel
NPS	nonpublic, nonsectarian school
NSLP	National School Lunch Program
PSAA	Public School Accountability Act
QC	quality control
RACF	Random Access Control Facility
SBE	State Board of Education
SD	standard deviation
SEM	standard error of measurement
SFTP	secure file transfer protocol
SGID	School and Grade Identification sheet
SMD	standardized mean difference
SPAR	Statewide Pupil Assessment Review
STAR	Standardized Testing and Reporting
STAR TAC	STAR Technical Assistance Center
STS	Standards-based Tests in Spanish
TIF	test information function
WRMSD	weighted root-mean-square differences

Chapter 1: Introduction

Background

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

In order to measure and evaluate student achievement of the content standards, the state instituted the Standardized Testing and Reporting (STAR) Program. This Program, administered annually, was authorized in 1997 by state law (Senate Bill 376).

During its 2011 administration, the STAR Program had four components:

- California Standards Tests (CSTs), produced for California public schools to assess the California content standards for ELA, mathematics, history-social science and science in grades two through eleven
- California Modified Assessment (CMA), an assessment of students' achievement of California's content standards for ELA, mathematics, and science, developed for students with an individualized education program (IEP) who meet the CMA eligibility criteria approved by the SBE
- California Alternate Performance Assessment (CAPA), produced for students with an IEP and who have significant cognitive disabilities and are not able to take the CSTs with accommodations and/or modifications or the CMA with accommodations
- Standards-based Tests in Spanish (STS), an assessment of students' achievement of California's content standards for Spanish-speaking English learners that is administered as the STAR Program's designated primary language test (DPLT)

Test Purpose

The CAPA is designed to show how well students with significant cognitive disabilities are performing with respect to California's content standards for ELA and mathematics in grades two through eleven and the content standards for science in grades five, eight, and ten. These standards describe what students should know and be able to do at each grade level. IEP teams determine on a student-by-student basis whether a student takes the CSTs, CMA, or the CAPA.

CAPA results are used in the school and district Academic Performance Index (API) calculations. In addition, CAPA results in grades two through eight and grade ten for ELA and mathematics are used in determining Adequate Yearly Progress (AYP), which applies toward meeting the requirement of the federal Elementary and Secondary Education Act (ESEA) that all students score at the proficient level or above by 2014.

Test Content

Students in grades two through eleven who take the CAPA are administered one of the five levels of the CAPA ELA and mathematics tests. In addition, students in grades five, eight, and ten take a grade-level science test.

The five levels of the CAPA are as follows:

- Level I, for students who are in grades two through eleven with the most significant cognitive disabilities
- Level II, for students who are in grades two and three
- Level III, for students who are in grades four and five
- Level IV, for students who are in grades six through eight
- Level V, for students who are in grades nine through eleven

Table 1.1 displays CAPA levels for tests administered in 2011 by grade, content area, and age ranges for ungraded programs.

Table 1.1 Description of the CAPA Assessment Levels

Test Level	I	II	III	IV	V
Grades	2–11	2 and 3	4 and 5	6–8	9–11
Content Area	ELA	ELA	ELA	ELA	ELA
	Mathematics	Mathematics	Mathematics	Mathematics	Mathematics
	Science Grades 5, 8, and 10 only	–	Science Grade 5 only	Science Grade 8 only	Science Grade 10 only
Age Ranges for Ungraded Programs	7–16	7 & 8	9 & 10	11–13	14–16

Intended Population

All students enrolled in grades two through eleven in California public schools on the day testing begins are required to take the CSTs, the CMA (available for students in grades three through eleven in ELA, grades three through seven in mathematics, EOC Algebra I and Geometry, and grades five, eight, and ten in science), or the CAPA. This requirement includes English learners regardless of the length of time they have been in U.S. schools or their fluency in English, as well as students with disabilities who receive special education services.

Students with significant cognitive disabilities and an IEP take the CAPA when they are unable to take the CSTs with or without accommodations and/or modifications or the CMA with accommodations. Most students eligible for the CAPA take the assessment level that corresponds with their current school grade, but some students with complex and profound disabilities take the Level I assessment. Level I is administered to students in grades two through eleven with the most significant cognitive disabilities who are receiving curriculum and instruction aligned to the CAPA Level I blueprints.

The decision to place a student in CAPA Level I must be made by the IEP team. Although it is possible that a student will take the CAPA Level I throughout his or her grade two through grade eleven education, the IEP team must reevaluate this decision each year. The decision to move a student from Level I to his or her grade-assigned CAPA level is made on the basis of both the student's CAPA performance from the previous year and on classroom assessments.

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

Intended Use and Purpose of Test Scores

The results for tests within the STAR Program are used for three primary purposes, described as follows (excerpted from the *EC* Section 60602 Web page at http://www.leginfo.ca.gov/cgi-bin/displaycode?section=edc&group=60001-61000&file=60600-60603) (outside source):

"60602. (a) (1) First and foremost, provide information on the academic status and progress of individual pupils to those pupils, their parents, and their teachers. This information should be designed to assist in the improvement of teaching and learning in California public classrooms. The Legislature recognizes that, in addition to statewide assessments that will occur as specified in this chapter, school districts will conduct additional ongoing pupil diagnostic assessment and provide information regarding pupil performance based on those assessments on a regular basis to parents or guardians and schools. The legislature further recognizes that local diagnostic assessment is a primary mechanism through which academic strengths and weaknesses are identified."

"60602. (a) (4) Provide information to pupils, parents or guardians, teachers, schools, and school districts on a timely basis so that the information can be used to further the development of the pupil and to improve the educational program."

"60602. (c) It is the intent of the Legislature that parents, classroom teachers, other educators, governing board members of school districts, and the public be involved, in an active and ongoing basis, in the design and implementation of the statewide pupil assessment program and the development of assessment instruments."

"60602. (d) It is the intent of the Legislature, insofar as is practically feasible and following the completion of annual testing, that the content, test structure, and test items in the assessments that are part of the Standardized Testing and Reporting Program become open and transparent to teachers, parents, and pupils, to assist all the stakeholders in working together to demonstrate improvement in pupil academic achievement. A planned change in annual test content, format, or design, should be made available to educators and the public well before the beginning of the school year in which the change will be implemented."

In addition, STAR Program assessments are used to provide data for school, district, and state purposes and to meet federal accountability requirements.

Testing Window

The CAPA is administered within a 21-day window which begins 10 days before and ends 10 days after the day on which 85 percent of the instructional year is completed. (This regulation was updated in June 2011. The new regulation, which will take effect beginning with the 2012 STAR administration, extends the STAR testing window by 4 days, from 21 instructional days to 25 instructional days.)

The CAPA tests are untimed. This assessment is administered individually and the testing time varies from one student to another, based on factors such as the student's response time and attention span. A student may be tested with the CAPA over as many days as required within the school district's testing window (California *Code of Regulations [CCR], Title 5,* Education, Division 1, Chapter 2, Subchapter 3.75, Article 2, § 855; in the California Department of Education [CDE] Web document at http://www.cde.ca.gov/ta/tg/sr/starregs0207cln.doc).

Significant STAR Developments in 2011

There were no new developments in the 2011 administration.

Limitations of the Assessment

Score Interpretation

Teachers and administrators should not use STAR results in isolation to make inferences about instructional needs. In addition, it is important to remember that a single test can provide only limited information. Other relevant information should be considered as well. It is advisable for parents to evaluate their child's strengths and weaknesses in the relevant topics by reviewing classroom work and progress reports in addition to the child's CAPA results (CDE, 2009). It is important to note that student scores in a content area contain measurement error and could vary if students were retested.

Out-of-Level Testing

With the exception of Level I, each CAPA is designed to measure the content corresponding to a specific grade or grade span, and is appropriate for students in the specific grade or grade span. Testing below a student's grade is not allowed for the CAPA or any test in the STAR Program; all students are required to take the test for the grade in which they are enrolled. School districts are advised to review all IEPs to ensure that any provision for testing below a student's grade level has been removed.

Score Comparison

When comparing results for the CAPA, the reviewer is limited to comparing results only within the same content area and CAPA level. For example, it is appropriate to compare scores obtained by students and/or schools on the 2011 CAPA Level II (Mathematics) test. Similarly, it is appropriate to compare scores obtained on the 2010 CAPA Level IV (ELA) test with those obtained on the CAPA Level IV (ELA) test administered in 2011. It is not appropriate to compare scores obtained on Levels II and IV of the ELA or mathematics tests, nor is it appropriate to compare ELA scores with mathematics scores. Since new score scales and cut scores were used for the 2009 CAPA tests, results from tests administered after 2009 cannot meaningfully be compared to results obtained in previous years.

Groups and Organizations Groups Involved in Test Development

State Board of Education

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

The SBE is responsible for assuring the compliance with programs that meet the requirement of the federal ESEA and the state's Public School Accountability Act (PSAA) and for reporting results in terms of the AYP and API, which measure the academic performance and growth of schools on a variety of academic measures. In order to provide information on student progress in public schools, as essential for those programs, the SBE supervises the administration and progress of the STAR Program.

California Department of Education

The CDE oversees California's public school system, which is responsible for the education of more than 7,000,000 children and young adults in more than 9,000 schools. California will provide a world-class education for all students, from early childhood to adulthood. The Department of Education serves California by innovating and collaborating with educators, schools, parents, and community partners which together, as a team, prepares students to live, work, and thrive in a highly connected world.

Contractors

Educational Testing Service

The CDE and the SBE contract with Educational Testing Service (ETS) to develop and administer the STAR Program. As the prime contractor, ETS has overall responsibility for working with the CDE to implement and maintain an effective assessment system and to coordinate the work of ETS and its subcontractor Pearson. Activities directly conducted by ETS include the following:

- Overall management of the program activities;
- Development of all test items;
- Construction and production of test booklets and related test materials;
- Support and training provided to counties, school districts, and independently testing charter schools;
- Implementation and maintenance of the STAR Management System for orders of materials and pre-identification services; and
- Completion of all psychometric activities.

Pearson

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

- Production of all scannable test materials;
- Packaging, distribution, and collection of testing materials to school districts and independently testing charter schools;
- Scanning and scoring of all responses, including performance scoring of the writing responses; and
- Production of all score reports and data files of test results.

Overview of the Technical Report

This technical report addresses the characteristics of the CAPA administered in spring 2011. The technical report contains nine additional chapters as follows:

- Chapter 2 presents a conceptual overview of processes involved in a testing cycle for a CAPA. This includes test construction, test administration, generation of test scores, and dissemination of score reports. Information about the distributions of scores aggregated by subgroups based on demographics and the use of special services is also included in this chapter. Also included are the references to various chapters that detail the processes briefly discussed in this chapter.
- Chapter 3 describes the procedures followed during the development of valid CAPA tasks; the chapter explains the process of field-testing new tasks and the review of tasks by contractors and content experts.

- Chapter 4 details the content and psychometric criteria that guided the construction of the CAPA for 2011.
- Chapter 5 presents the processes involved in the actual administration of the 2011 CAPA with an emphasis on efforts made to ensure standardization of the tests. It also includes a detailed section that describes the procedures that were followed by ETS to ensure test security.
- Chapter 6 describes the standard-setting process conducted to establish new cut scores.
- Chapter 7 details the types of scores and score reports that are produced at the end of each administration of the CAPA.
- Chapter 8 summarizes the results of the task (item)-level analyses performed during the spring 2011 administration of the tests. These include the classical item analyses, the reliability analyses that include assessments of test reliability and the consistency and accuracy of the CAPA performance-level classifications, and the procedures designed to endure the validity of CAPA score uses and interpretations. Also discussed in this chapter are the item response theory (IRT) and model-fit analyses, as well as documentation of the equating along with CAPA conversion tables. Finally, the chapter summarizes the results of analyses investigating the differential item functioning (DIF) for each CAPA.
- Chapter 9 highlights the importance of controlling and maintaining the quality of the CAPA.
- Chapter 10 presents historical comparisons of various task (item)- and test-level results for the past two years and for the 2009 base year.

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

Reference

California Department of Education. (2009). *Interpreting 2009 STAR program test results*. Sacramento, CA.

Chapter 2: An Overview of CAPA Processes

This chapter provides an overview of the processes involved in a typical test development and administration cycle for the CAPA. Also described are the specifications maintained by ETS to carry out each of those processes. The chapter is organized to provide a brief description of each process followed by a summary of the associated specifications. More details about the specifications and the analyses associated with each process are described in other chapters that are referenced in the sections that follow.

Task (Item) Development

Task Formats

Each CAPA task involves a prompt that asks a student to perform a task or a series of tasks. Each CAPA task consists of the Task Preparation, the Cue/Direction, and the Scoring Rubrics. The rubrics define the rules for scoring a student's response to each task.

Task (Item) Specifications

The CAPA tasks are developed to measure California content standards and designed to conform to principles of task writing defined by ETS (ETS, 2002). ETS maintains task and updates a task specifications document, otherwise known as "task writer guidelines," for each CAPA and has developed an item utilization plan to guide the development of the tasks for each content area. Task writing emphasis is determined in consultation with the CDE.

The task specifications describe the characteristics of the tasks that should be written to measure each content standard. The task specifications help ensure that the tasks in the CAPA measure the content standards in the same way. To do this, the task specifications provide detailed information to task writers that are developing tasks for the CAPA.

The tasks selected for each CAPA undergo an extensive review process that is designed to provide the best standards-based tests possible. Details about the task specifications, the task review process, and the item utilization plan are presented in Chapter 3, starting on page 16.

Item Banking

Before newly developed tasks are placed in the item bank, ETS prepares the tasks for review by content experts and various external review organizations such as the Assessment Review Panels (ARPs), which are described in Chapter 3 starting on page 20; and the Statewide Pupil Assessment Review (SPAR) panel, described in Chapter 3 starting on page 22.

Once the ARP review is complete, the tasks are then placed in the item bank along with their corresponding information obtained at the review sessions. Tasks that are accepted by the content experts are updated to a "field-test ready" status; tasks that are rejected are updated to a "rejected before use" status. ETS then delivers the tasks to the CDE by means of a delivery of the California electronic item bank. Tasks are then field tested to obtain information about item performance and item statistics that can be used to assemble operational forms. Subsequent updates to item content and statistics are based on data collected from the operational use of the items. However, only the latest content of the item is retained in the bank at any time, along with the administration data from every administration that has included the item.

Further details on item banking are presented on page 24 in Chapter 3.

Task Refresh Rate

The item utilization plan assumes that each year, 30 percent of tasks on an operational form are refreshed (replaced); these tasks remain in the item bank for future use.

Test Assembly

Test Length

Each CAPA consists of twelve tasks, including eight operational tasks and four field-test tasks. The number of tasks in each CAPA and the expected time to complete a test is presented in Table 2.1 Testing times for the CAPA are approximate. This assessment is administered individually and the testing time varies from one student to another based on factors such as the student's response time and attention span. A student may be tested with the CAPA over as many days as required within the school district's selected testing window.

Table 2.1 CAPA Item and Estimated Time Chart

ITEM and ESTIMATED TIME CHART		
	Grades 2–11	
CAPA Content Area	Items	Times
English–Language Arts	12	45 minutes
Mathematics	12	45 minutes
Science	12	45 minutes

Test Blueprints

ETS selects all CAPA tasks to conform to the SBE-approved California content standards and test blueprints. The CAPA has been revised to better link it to the grade-level California content standards. The revised blueprints for the CAPA were approved by the SBE in 2006 for implementation beginning in 2008. The test blueprints for the CAPA can be found on the CDE STAR CAPA Blueprints Web page at http://www.cde.ca.gov/ta/tg/sr/capablueprints.asp.

Content Rules and Task Selection

When developing a new test form for a given CAPA level and content area, test developers follow a number of rules. First and foremost, they select tasks that meet the blueprint for that level and content area. Using the electronic item bank, assessment specialists begin by identifying a number of linking tasks. These are tasks that appeared in the previous year's operational administration and are used to equate the test forms administered each year. Linking tasks are selected to proportionally represent the full blueprint. Each CAPA form is a collection of test tasks designed to reflect a reliable, fair, and valid measure of student learning within well-defined course content.

Another consideration is the difficulty of each task. Test developers strive to ensure that there are some easy and some hard tasks and that there are a number of tasks in the middle range of difficulty. The detailed rules are presented in Chapter 4, which begins on page 26.

Psychometric Criteria

The staff assesses the projected test characteristics during the preliminary review of the assembled forms. The statistical targets used for the 2011 test development and the

projected characteristics of the assembled forms are presented starting from page 27 in Chapter 4.

The tasks in test forms are organized and sequenced to meet the requirements of the content area. Further details on the arrangement of tasks during test assembly are also described on page 28 in Chapter 4.

Test Administration

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

Test Security and Confidentiality

All tests within the STAR Program are secure documents. For the CAPA administration, every person having access to test materials maintains the security and confidentiality of the tests. ETS's Code of Ethics requires that all test information, including tangible materials (such as test booklets, test questions, test results), confidential files, processes, and activities are kept secure. To ensure security for all tests that ETS develops or handles, ETS maintains an Office of Testing Integrity (OTI). A detailed description of the OTI and its mission is presented in Chapter 5 on page 29.

In its pursuit of enforcing secure practices, ETS and the OTI strive to safeguard the various processes involved in a test development and administration cycle. Those processes are listed below. The practices related to each process are discussed in detail in Chapter 5, starting on page 29, and are as follows:

- Test development
- Task and data review
- Item banking
- Transfer of forms and tasks to the CDE
- Security of electronic files using a firewall
- Printing and publishing
- Test administration
- Test delivery
- Processing and scoring
- Data management
- Transfer of scores via secure data exchange
- Statistical analysis
- Reporting and posting results
- Student confidentiality
- Student test results

Procedures to Maintain Standardization

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

Test Administrators

The CAPA is administered in conjunction with the other tests that comprise the STAR Program. ETS employs personnel who facilitate various processes involved in the standardization of an administration cycle.

Staff at school districts who are central to the processes include district STAR coordinators, test site coordinators, test examiners, test proctors, and observers. The responsibilities for each of the staff members are included in the *STAR District and Test Site Coordinator Manua*l (CDE, 2011a); see page 35 in Chapter 5 for more information.

Test Directions

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

***CAPA Examiner's Manual*—**The manual used by test examiners to administer and score the CAPA to be followed exactly so that all students have an equal opportunity to demonstrate their academic achievement (See page 35 in Chapter 5 for more information.)

***District and Test Site Coordinator Manual*—**Test administration procedures for district STAR coordinators and test site coordinators (See page 35 in Chapter 5 for more information.)

STAR Management System manuals—Instructions for the Web-based modules that allow district STAR coordinators to set up test administrations, order materials, and submit and correct student Pre-ID data; every module has its own user manual with detailed instructions on how to use the STAR Management System (See page 36 in Chapter 5 for more information.)

Test Variations, Accommodations, and Modifications

All public school students participate in the STAR Program, including English learners and students with disabilities. Students with an IEP and who have significant cognitive disabilities may take the CAPA when they are unable to take the CSTs with or without accommodations and/or modifications or the CMA with accommodations.

Examiners may adapt the CAPA in light of a student's instructional mode as specified in each student's IEP or Section 504 plan in one of two ways (1) Suggested adaptations for particular tasks, as specified in the task preparation; and (2) Core adaptations that are applicable for many of the tasks. Details of the adaptations are presented in the core adaptations of the *CAPA Examiner's Manual* (CDE, 2011b).

As noted on the CDE CAPA Participation Criteria Web page, "Since examiners may adapt the CAPA based on students' instructional mode, accommodations and modifications do not apply to the CAPA." (CDE, 2011c)

Scores

The CAPA total raw scores equal the sum of examinees' scores on operational tasks. Raw scores for Level I range from 0 to 40; for the other CAPA levels, the raw-score range is from 0 to 32. Those raw scores are transformed to two-digit scale scores using the scaling process described starting on page 13. CAPA results are reported through the use of these scale scores; the scores range from 15 to 60 for each test. Also reported are performance levels obtained by classifying the scale scores into the following levels: far below basic, below basic, basic, proficient, and advanced. The state's target is for all students to score at the proficient or advanced level.

Detailed descriptions of CAPA scores are described on page 45 in Chapter 7.

Aggregation Procedures

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

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

1. Met attemptedness criteria
2. A valid combination of grade and CAPA level
3. Did not have a parental exemption

Individual Scores

Table 7.2 through Table 7.4 starting on page 48 in Chapter 7 provide summary statistics for individual scores aggregated at the state level, describing overall student performance on each CAPA. Included in the tables are the possible and actual ranges, and the means and standard deviations of student scores, expressed in terms of both raw scores and scale scores. The tables also present statistical information about the CAPA tasks.

Demographic Subgroup Scores

Statistics summarizing CAPA student performance by content area and for selected groups of students are provided in Table 7.B.1 through Table 7.B.3 starting on page 59 in Appendix 7B. In these tables, students are grouped by demographic characteristics, including gender, ethnicity, English-language fluency, primary disability, and economic status. The tables show the numbers of students with valid scores in each group, scale score means and standard deviations as well as percentage in performance level for each demographic group. Table 7.6 on page 52 provides definitions for the demographic groups included in the tables.

Equating

Each CAPA is equated to a reference form using a common-item nonequivalent groups data collection design and methods based on item response theory (IRT). The “base” or “reference” calibrations for the CAPA were established by calibrating samples of data from the 2009 administration. Doing so established a scale to which subsequent item calibrations could be linked. The 2011 items were placed on the reference 2009 scale using a set of linking items selected from the 2009 forms and re-administrated in 2011.

The procedure used for equating the CAPA involves three steps: calibration, scaling, and linear transformation. Each of those procedures, as described below, is applied to all CAPA tests.

Calibration

To obtain item calibrations, a proprietary version of the PARSCALE program and Rasch partial credit model are used. The estimation process is constrained by setting a common discrimination value for all tasks equal to 1.0 / 1.7 (or 0.588). This approach is in keeping with previous CAPA calibration procedures accomplished using the WINSTEPS program (Linacre, 2000).

The PARSCALE calibrations are run in two stages, following procedures used with other ETS testing programs. In the first stage, estimation imposed normal constraints on the updated prior ability distribution. The estimates resulting from this first stage 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.

Scaling

Calibrations of the 2011 tasks were linked to the previously obtained reference scale estimates using linking tasks and the Stocking and Lord (1983) procedure. In the case of the one-parameter model calibrations, this procedure is equivalent to setting the mean of the new task parameter estimates for the linking set equal to the mean of the previously scaled estimates. As noted earlier, the linking set is a collection of tasks in a current test form that also appeared in last year's form and was scaled at that time.

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

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

where,

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

n_g is the number of intervals/groups,

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

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

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

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

Based on established procedures, any linking items for which the WRMSD was greater than 0.625 for Level I and 0.500 for Levels II through V were eliminated. This criterion has produced reasonable results over time in similar equating work done with other testing programs at ETS.

Linear Transformation

Once the new task calibrations for each test were transformed to the base scale, raw score to theta scoring tables were generated. The thetas in these tables were then linearly transformed to a two-digit score scale that ranged from 15 to 60. Because the basic and proficiency cut scores were required to be equal to 30 and 35, respectively, the following formula was used to make this transformation:

$$ScaleScore = \left(35 - \theta_{pro} \times \left(\frac{35-30}{\theta_{pro} - \theta_{bas}} \right) \right) + \left(\frac{35-30}{\theta_{pro} - \theta_{bas}} \right) \times \theta \tag{2.2}$$

where,

θ represents student ability,

θ_{pro} represents theta cut score for proficient on spring 2009 base scale, and

θ_{bas} represents theta cut score for basic on spring 2009 base scale.

Complete raw-to-scale score conversion tables for the 2011 CAPA are presented in Table 8.D.10 through Table 8.D.23 in Appendix 8.D, starting on page 140. The raw scores and corresponding rounded converted scale scores are listed in those tables.

The scale scores defining the various performance levels are presented in Table 2.2.

Table 2.2 Scale Scores Ranges for Performance Levels

Content Area	CAPA Level	Far Below Basic	Below Basic	Basic	Proficient	Advanced
English–Language Arts	I	15	16 – 29	30 – 34	35 – 39	40 – 60
	II	15 – 18	19 – 29	30 – 34	35 – 39	40 – 60
	III	15 – 23	24 – 29	30 – 34	35 – 39	40 – 60
	IV	15 – 17	18 – 29	30 – 34	35 – 41	42 – 60
	V	15 – 22	23 – 29	30 – 34	35 – 39	40 – 60
Mathematics	I	15	16 – 29	30 – 34	35 – 38	39 – 60
	II	15 – 17	18 – 29	30 – 34	35 – 40	41 – 60
	III	15	16 – 29	30 – 34	35 – 39	40 – 60
	IV	15	16 – 29	30 – 34	35 – 40	41 – 60
	V	15 – 16	17 – 29	30 – 34	35 – 39	40 – 60
Science	I	15	16 – 29	30 – 34	35 – 38	39 – 60
	III	15 – 21	22 – 29	30 – 34	35 – 39	40 – 60
	IV	15 – 19	20 – 29	30 – 34	35 – 39	40 – 60
	V	15 – 20	21 – 29	30 – 34	35 – 38	39 – 60

References

California Department of Education. (2011a). *2011 STAR district and test site coordinator manual*. Sacramento, CA. Downloaded from http://www.startest.org/pdfs/STAR.coord_man.2011.pdf

California Department of Education. (2011b). *2011 California Alternate Performance Assessment (CAPA) Examiner's Manual*. Sacramento, CA. Downloaded from http://www.startest.org/pdfs/CAPA.examiners_manual.nonsecure.2011.pdf

California Department of Education. (2011c). *CAPA participation criteria.* Downloaded from http://www.cde.ca.gov/TA/tg/sr/participcritria.asp

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

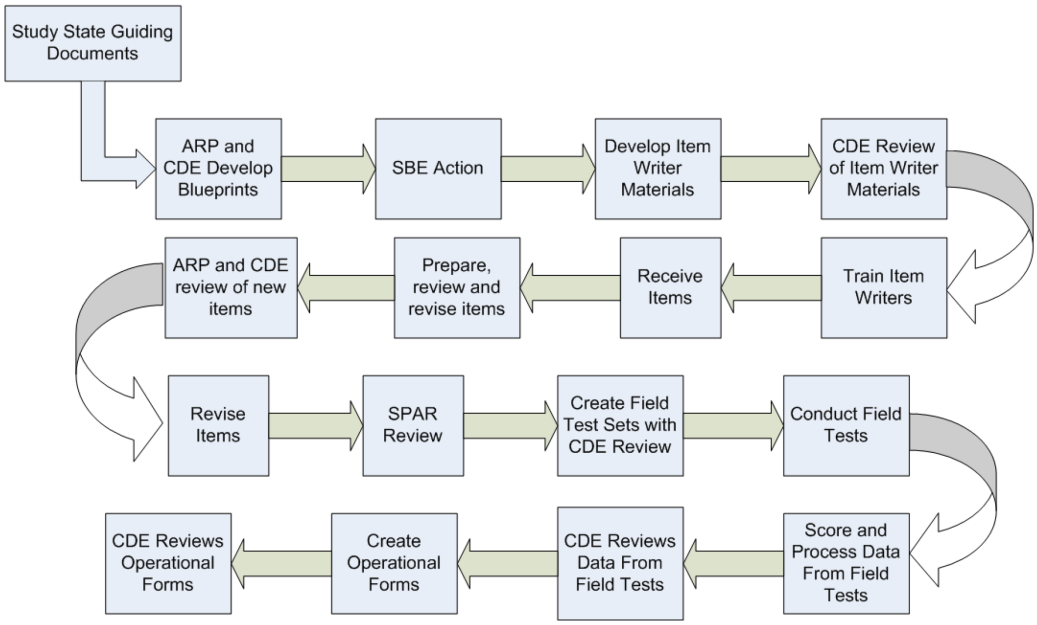
Linacre, J. M. (2000). *WINSTEPS: Rasch measurement* (Version 3.23). Chicago, IL: MESA Press.

Stocking, M. L., and Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*, pp. 201–10.

Chapter 3: Task (Item) Development

The CAPA tasks are developed to measure California's content standards and designed to conform to principles of item writing defined by ETS (ETS, 2002). Each CAPA task goes through a comprehensive development cycle as is described in Figure 3.1, below.

Figure 3.1 The ETS Item Development Process for the STAR Program



Rules for Task Development

The development of CAPA tasks follow guidelines for task writing approved by the CDE. These guidelines direct a task writer to assess a task for the relevance of the information being assessed, its relevance to the California content standards, its match to the test and task specifications, and its appropriateness to the population being assessed. As described below, tasks are eliminated early in a rigorous task review process when they are only peripherally related to the test and task specifications, do not measure core outcomes reflected in the California content standards, or are not developmentally appropriate.

Task Specifications

ETS senior content staff leads the task writers in the task development and review process. In addition, experienced ETS content specialists and assessment editors review each task during the forms-construction process. The lead assessment specialists for each content area work directly with the other ETS assessment specialists to carefully review and edit each task for such technical characteristics as quality, match to content standards, and conformity with California-approved task-writing practices. ETS follows the SBE-approved item utilization plan to guide the development of the tasks for each content area. Task specification documents include a description of the constructs to be measured and the California content standards. Those specifications help to ensure that the CAPA measures the content standards in the same way each year. The task specifications also provide specific and important guidance to task writers.

The task specifications describe the general characteristics of the tasks for each content standard, indicate task types or content to be avoided, and define the content limits for the tasks. More specifically, 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 CAPA blueprint
- The construct(s) appropriately measured by the standard
- A description of specific kinds of tasks to be avoided, if any (such as ELA tasks about insignificant details)
- A description of appropriate data representations (such as charts, tables, graphs, or other artwork) for mathematics and science tasks
- The content limits for the standard (such as one or two variables, maximum place values of numbers) for mathematics and science tasks
- A description of appropriate stimulus cards (if applicable) for ELA tasks

In addition, the ELA task specifications that contain guidelines for stimulus cards used to assess reading comprehension include the following:

- A list of topics to be avoided
- The acceptable ranges for the number of words on a stimulus card
- Expected use of artwork
- The target number of tasks attached to each reading stimulus card

Expected Task Ratio

ETS has developed the item utilization plan to continue the development of CAPA tasks. The plan includes strategies for developing tasks that will permit coverage of all appropriate standards for all tests in each content area and at each grade level. ETS test development staff uses this plan to determine the number of tasks to develop for each content area.

The item utilization plan assumes that each year, 35 percent of items on an operational form would be refreshed (replaced); these items remain in the item bank for future use.

The item utilization plan also declares that an additional five percent of the operational items are likely to become unusable because of normal attrition, and notes that there is a need to focus development on “critical” standards, which are standards that are difficult to measure well or for which we have few usable items.

Each year, ETS field tests 16 tasks per CAPA level for both ELA and mathematics, and eight tasks per CAPA level for science. Given that each test contains eight operational tasks, the ratios of field-test to operational tasks are 200 percent for ELA and mathematics and 100 percent for science for each CAPA level. These task ratios would allow for a five percent attrition rate while gradually increasing the overall size of the CAPA item bank. The field-test percentages and task counts are presented in Table 3.1.

Table 3.1 Field-test Percentages for the CAPA

Content Area	Number of Operational Tasks per CAPA level	Field-test Percentage per CAPA level	Number of Tasks to Be Field-tested per CAPA level
English–Language Arts	8	200%	16
Mathematics	8	200%	16
Science	8	100%	8

Selection of Task Writers

Criteria for Selecting Task Writers

The tasks for each CAPA are written by individual task writers who have a thorough understanding of the California content standards. Applicants for task writing are screened by senior ETS content staff. Only those with strong content and teaching backgrounds are approved for inclusion in the training program for task writers. Because most of the participants are current or former California educators, they are particularly knowledgeable about the standards assessed by the CAPA. All task writers meet the following minimum qualifications:

- Possession of a bachelor's degree in the relevant content area or in the field of education with special focus on a particular content of interest; an advanced degree in the relevant content area is desirable
- Previous experience in writing tasks for standards-based assessments, including knowledge of the many considerations that are important when developing tasks to measure state-specific standards
- Previous experience in writing tasks in the content areas covered by CAPA levels
- Familiarity, understanding, and support of the California content standards
- Current or previous teaching experience in California, when possible
- Knowledge about the abilities of the students taking the tests

Task (Item) Review Process

The tasks selected for CAPA undergo an extensive task review process that is designed to provide the best standards-based tests possible. This section summarizes the various reviews performed to ensure the quality of the CAPA tasks and test forms.

Contractor Review

Once the tasks have been written, ETS employs a series of internal reviews. The reviews establish the criteria used to judge the quality of the task content and are designed to ensure that each task is measuring what it is intended to measure. The internal reviews also examine the overall quality of the tasks before they are prepared for presentation to the CDE and the Assessment Review Panels (ARPs). Because of the complexities involved in producing defensible tasks for high-stakes programs such as the STAR Program, it is essential that many experienced individuals review each task before it is brought to the CDE, the ARPs and Statewide Pupil Assessment Review (SPAR) panels.

The ETS review process for the CAPA includes the following:

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

Throughout this multistep task review process, the lead content-area assessment specialists and development team members continually evaluate the relevance of the information being assessed by the task, its relevance to the California content standards, its match to the test and task specifications, and its appropriateness to the population being assessed. Tasks that are only peripherally related to the test and task 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.

1. Internal Content Review

Test tasks and materials undergo two reviews by the content area assessment specialists. These assessment specialists make sure that the test tasks and related materials are in compliance with ETS's written guidelines for clarity, style, accuracy, and appropriateness for California students as well as in compliance with the approved task specifications. Assessment specialists review each task on the basis of the following characteristics:

- Relevance of each task as the task relates to the purpose of the test
- Match of each task to the task specifications, including cognitive level
- Match of each task to the principles of quality task development
- Match of each task to the identified standard or standards
- Difficulty of the task
- Accuracy of the content of the task
- Readability of the task or stimulus card
- CAPA-level appropriateness of the task
- Appropriateness of any illustrations, graphs, or figures

Each task is classified with a code for the standard it is intended to measure. The assessment specialists check all tasks 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 may accept the task and classification as written, suggest revisions, or recommend that the task be discarded. These steps occur prior to the CDE's review.

2. Internal Editorial Review

After the content area assessment specialists review each task, a group of specially trained editors reviews each task in preparation for review by the CDE and the ARPs. The editors check tasks for clarity, correctness of language, appropriateness of language for the grade level assessed, adherence to the style guidelines, and conformity with accepted task-writing practices.

3. Internal Sensitivity Review

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

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
- Task accessibility for English-language learners

Content Expert Reviews

Assessment Review Panels

ETS is responsible for working with ARPs as tasks are developed for the CAPA. The ARPs are advisory panels to the CDE and ETS and provide guidance on matters related to task development for the CAPA. The ARPs are responsible for reviewing all newly developed tasks for alignment to the California content standards. The ARPs also review the tasks for accuracy of content, clarity of phrasing, and quality. In their examination of test tasks, the ARPs may raise concerns related to age/level appropriateness and gender, racial, ethnic, and/or socioeconomic bias.

Composition of ARPs

The ARPs are composed 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 CAPA ARPs, including:

- Three or more years of general teaching experience in grades kindergarten through twelve and in the content areas (ELA, mathematics, or science);
- Bachelor's or higher degree in a grades or content area related to ELA, mathematics, or science;
- Knowledge and experience with the California content standards for ELA, mathematics, or science;
- Special education credential;
- Experience with more than one type of disability; and
- Three to five years as a teacher or school administrator with a special education credential.

Every effort is made to ensure that the 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 are recruited through an application process. Recommendations are solicited from school districts and county offices of education as well as from CDE and SBE staff. Applications are received and reviewed throughout the year. They are reviewed by the ETS assessment directors, who confirm that the applicant's qualifications meet the specified criteria. Applications that meet the criteria are forwarded to CDE and SBE staff for further review and agreement on ARP membership. Upon approval, the applicant is notified that he or she has been selected to serve on the ARP committee.

Table 3.2, on the next page, shows the educational qualifications, present occupation, and credentials of the current CAPA ARP members.

Table 3.2 CAPA ARP Member Qualifications, by Content Area and Total

CAPA	ELA	Math	Science	Total
Total	7	8	5	20
Occupation (Members may teach multiple levels.)				
Teacher or Program Specialist, Elementary/Middle School	4	2	1	7
Teacher or Program Specialist, High School	1	1	2	4
Teacher or Program Specialist, K–12	4	3	4	11
University Personnel	0	0	0	0
Other District Personnel (e.g., Director of Special Services, etc.)	2	1	0	3
Highest Degree Earned				
Bachelor's Degree	4	4	2	10
Master's Degree	3	4	3	10
Doctorate	0	0	0	0
K–12 Teaching Credentials and Experience (Members may hold multiple credentials.)				
Elementary Teaching (multiple subjects)	4	3	1	8
Secondary Teaching (single subject)	0	1	4	5
Special Education	4	6	4	14
Reading Specialist	0	0	0	0
English Learner (CLAD,BCLAD)	1	1	1	3
Administrative	1	1	2	4
Other	0	0	0	0
None (teaching at the university level)	0	0	0	0

ARP Meetings for Review of CAPA Tasks

ETS content-area assessment specialists facilitate the CAPA ARP meetings. Each meeting begins with a brief training session on how to review tasks. ETS provides this training, which consists of the following topics:

- Overview of the purpose and scope of the CAPA
- Overview of the CAPA's test design specifications and blueprints
- Analysis of the CAPA's task specifications
- Overview of criteria for reviewing constructed-response tasks
- Review and evaluation of tasks for bias and sensitivity issues

Criteria also involve more global factors, including—for ELA—the appropriateness, difficulty, and readability of reading stimulus cards. The ARPs also are trained on how to make recommendations for revising tasks.

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

Does the task:

- Measure the content standard?
- Match the test task specifications?
- Align with the construct being measured?

- Test worthwhile concepts or information?
- Reflect good and current teaching practices?
- Have wording that gives the student a full sense of what the task is asking?
- Avoid unnecessary wordiness?
- Reflect content that is free of bias against any person or group?

Is the stimulus, if any, for the task:

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

As the first step of the task review process, ARP members review a set of tasks independently and record their individual comments. The next step in the review process is for the group to discuss each task. The content-area assessment specialists facilitate the discussion and record all recommendations. Those recommendations are recorded in a master task-review booklet. Task review binders and other task evaluation materials also serve to identify potential bias and sensitivity factors that the ARP will consider as a part of its task reviews.

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

Statewide Pupil Assessment Review Panel

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

Field Testing

The primary purposes of field testing are to obtain information about task performance and to obtain statistics that can be used to assemble operational forms.

Stand-alone Field Testing

In 2002, for the new CAPA, a pool of tasks was initially constructed by administering the newly developed tasks in a stand-alone field test. In stand-alone field testing, examinees are recruited to take tests outside of the usual testing situation and the test results are typically not used for instructional or accountability purposes (Schmeiser & Welch, 2006).

Embedded Field-test Tasks

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

field-tested tasks are later administered operationally. Such field testing involves distributing the tasks being field-tested within an operational test form. Different forms contain the same operational tasks and different field-test tasks. The numbers of embedded field-test tasks for the CAPA are shown in Table 3.3.

Allocation of Students to Forms

The test forms for a given CAPA are distributed by random assignment to school districts and independently testing charter schools so that a large representative sample of test takers responds to the field-test items embedded in these forms. The random assignment of specific forms ensures that a diverse sample of students take each field-test task.

The students do not know which tasks are field-test tasks and which tasks are operational tasks; therefore, their motivation is not expected to vary over the two types of tasks (Patrick & Way, 2008).

Number of Forms and Sample Sizes

All CAPA assessments consist of four forms. Each form contains eight operational tasks that are the same and four unique tasks being field-tested. Scores on the field-test tasks are not counted toward student scores. See Table 2.1 on page 9 for more details on the test length.

Table 3.3 shows the number of forms, operational tasks, field-test tasks, and the approximate number of students in the P2 sample that took the operational and field-test tasks in spring 2011. The P2 data file contains 100 percent of school district data that were received for ETS statistical analysis by approximately August 23, 2011.

The sample sizes for the field-test tasks are presented as ranges because the numbers of students who took a set of field-test tasks varied over the forms of CAPA.

Table 3.3 Summary of Tasks and Forms Presented in the 2011 CAPA

Content Area	Level	Operational		Field Test		
		No. Tasks	Examinees Total (P2)	No. Forms	No. Tasks	Examinees Total (P2)
English–Language Arts	I	8	13,719	4	16	2,741–3,821
	II	8	6,643	4	16	1,199–1,743
	III	8	7,112	4	16	1,296–1,858
	IV	8	9,858	4	16	1,864–2,633
	V	8	10,217	4	16	1,818–2,578
Mathematics	I	8	13,689	4	16	2,732–3,815
	II	8	6,624	4	16	1,196–1,739
	III	8	7,098	4	16	1,293–1,854
	IV	8	9,845	4	16	1,858–2,629
	V	8	10,196	4	16	1,814–2,577
Science	I	8	3,512	4*	8	646–1,032
	III	8	3,391	4*	8	600–893
	IV	8	3,155	4*	8	630–799
	V	8	3,245	4*	8	594–788

* There are two unique forms and two repeated forms for science tests.

CDE Data Review

Once tasks have been field tested, ETS prepares the tasks and the associated statistics for review by the CDE. ETS provides tasks with their statistical data, along with annotated comments sheets, for the CDE to use in its review. ETS conducts an introductory training to highlight any new issues and serve as a statistical refresher. CDE consultants then make decisions about which tasks should be included in the item bank. ETS psychometric and content staffs are available to CDE consultants throughout this process.

Item Banking

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

After field-test or operational use, tasks that do not meet statistical specification may be rejected; such tasks are updated with a status of “rejected for statistical reasons” and remain unavailable in the bank. These statistics are obtained by the psychometrics group at ETS, which carefully evaluates each task for its level of difficulty and discrimination as well as conformance to the IRT model. Psychometricians also determine if the task functions similarly for various subgroups of interest.

Status and availability of a task are updated programmatically as tasks are presented for review, accepted or rejected, placed on a form for field-testing, presented for statistical review, and used operationally. All rejection indications are monitored and controlled through ETS’s assessment development processes.

ETS currently provides and maintains the electronic item banks for several of the California assessments including the California High School Exit Examination (CAHSEE) and STAR (CST, CMA, CAPA, and STS). CAHSEE and STAR are currently consolidated in the California Item Banking system. ETS works with the CDE to obtain the data for assessments under contract with other vendors for inclusion into the item bank. ETS provides the item banking application using the LAN architecture and the relational database management system, SQL 2000, already deployed. ETS provides updated versions of the item bank to the CDE on an ongoing basis and works with the CDE to determine the optimum process if a change in databases is desired.

References

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

Patrick, R., & Way, D. (March, 2008). *Field testing and equating designs for state educational assessments*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Schmeiser, C. B., & Welch, C. J. (2006). Test development. In R.L. Brennan (Ed.), *Educational measurement* (4th, ed.). Westport, CT: American Council on Education and Praeger Publishers.

Chapter 4: Test Assembly

The CAPA is constructed to measure students' performance relative to California's content standards approved by the SBE. The tests are also constructed to meet professional standards for validity and reliability. For each CAPA, the content standards and desired psychometric attributes are used as the basis for assembling the test forms.

Test Length

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

Each CAPA consists of 12 tasks, including eight operational tasks and four field-test tasks. For more details on the distribution of items at each level and content area, see Table 3.3 in Chapter 3 on page 23.

Rules for Task Selection

Test Blueprints

ETS develops all CAPA test tasks to conform to the SBE-approved California content standards and test blueprints. The CAPA blueprints were revised and approved by the SBE in 2006 for implementation beginning in 2008.

The California content standards were used as the basis for choosing tasks for the tests. The blueprints for the CAPA can be found on the following CDE STAR CAPA Blueprints Web page at http://www.cde.ca.gov/ta/tg/sr/capablueprints.asp.

Content Rules and Task Selection

When developing a new test form for a given CAPA level and content area, test developers follow a number of rules. First and foremost, they select tasks that meet the blueprint for that level and content area. Using an electronic item bank, assessment specialists begin by identifying a number of linking tasks. These are tasks that appeared in the previous year's operational administration and are used to equate the test forms administered each year. Linking tasks are selected to proportionally represent the full blueprint. The selected linking items are also reviewed by psychometricians to ensure that the specific psychometric criteria are met.

After the linking tasks are approved, assessment specialists populate the rest of the test form. Their first consideration is the strength of the content and the match of each task to a content standard. In selecting tasks, team members also try to ensure that they include a variety of formats and content and that at least some of them include graphics for visual interest.

Another consideration is the difficulty of each task. Test developers strive to ensure that there are some easy and some hard tasks, and that there are a number of tasks in the middle range of difficulty. If tasks do not meet all content and psychometric criteria, staff reviews the other available tasks to determine if there are other selections that could improve the match of the test to all of the requirements. If such a match is not attainable, the content team works in conjunction with psychometricians and the CDE to determine which

combination of tasks will best serve the needs of the students taking the test. Chapter 3 on page 16 contains further information about this process.

Psychometric Criteria

The three goals of CAPA test development are:

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

In order to achieve these goals, a set of rules that outlines the desired psychometric properties of the CAPA has been developed. These rules are referred to as statistical targets. Total test assembly targets are developed for each CAPA. These targets are provided to test developers before a test construction cycle begins.

The total test targets, or primary statistical target, used for assembling the CAPA forms for the 2011 administration were the average and standard deviation of item difficulty based on the item response theory (IRT) *b*-parameters, average item score (AIS), and average polyserial correlation. Due to the unique characteristics of the Rasch IRT model, the information curve conditional on each ability level is determined by item difficulty alone. Therefore, the test information function (TIF) suffices as the target for conditional test difficulty. Although additional item difficulty targets are not imperative when the target TIF is used for form construction, the target mean and standard deviation of item difficulty consistent with the TIF were also provided to test development staff to help with the test construction process.

The polyserial correlation describes the relationship between student performance on a polytomously scored item and student performance on the test as a whole. It is used as a measure of how well an item discriminates among test takers that differ in their ability and is related to the overall reliability of the test.

The target values for the CAPA, presented in Table 4.1, were those used to build the spring 2011 operational test forms. These specifications were developed from the analyses of test forms administered in 2009, the base year in which test results were reported using new scale and cut scores for the five performance levels: far below basic, below basic, basic, proficient, and advanced.

Table 4.1 Target Statistical Specifications for the CAPA

Content Area	CAPA Level	Target Mean *b*	Target SD *b*	Mean AIS	Mean Polyserial
English–Language Arts	I	–0.39	0.50	2.75	0.80
	II	–0.56	0.50	2.20	0.80
	III	–0.49	0.50	2.20	0.80
	IV	–0.50	0.50	2.20	0.80
	V	–0.61	0.50	2.20	0.80
Mathematics	I	–0.27	0.50	2.75	0.80
	II	–0.79	0.50	2.20	0.80
	III	–0.80	0.50	2.20	0.80
	IV	–0.73	0.50	2.20	0.80
	V	–0.79	0.50	2.20	0.80

Content Area	CAPA Level	Target Mean *b*	Target SD *b*	Mean AIS	Mean Polyserial
Science	I	–0.27	0.50	2.75	0.80
	III	–0.76	0.50	2.20	0.80
	IV	–0.61	0.50	2.20	0.80
	V	–0.31	0.50	2.20	0.80

Projected Psychometric Properties of the Assembled Tests

Prior to the 2011 administration, ETS psychometricians performed a preliminary review of the technical characteristics of the assembled tests. The expected or projected performance of examinees was estimated using the item level statistics, available in the California item bank, for the selected items. Table 4.2 shows the mean observed statistics of the items on each CAPA, based on the most recent administration of those items. These values can be compared to the target values in Table 4.1.

Table 4.2 Summary of 2011 CAPA Projected Statistical Attributes

Content Area	CAPA Level	Mean *b*	SD *b*	Mean AIS	Min AIS	Max AIS	Mean Polyserial
English–Language Arts	I	–0.62	0.22	3.19	2.73	4.12	0.76
	II	–1.01	0.54	2.47	1.90	3.53	0.70
	III	–1.29	0.79	2.85	2.21	3.60	0.75
	IV	–1.07	0.67	2.56	1.66	3.51	0.73
	V	–1.10	0.56	2.56	0.77	3.40	0.93
Mathematics	I	–0.22	0.14	2.79	2.52	3.20	0.73
	II	–0.91	0.76	2.53	1.42	3.48	0.70
	III	–0.69	0.58	2.32	1.72	2.97	0.63
	IV	–0.82	0.63	2.47	1.83	3.02	0.74
	V	–0.96	0.33	2.51	2.03	3.18	0.74
Science	I	–0.33	0.14	2.92	2.60	3.51	0.79
	III	–1.11	0.26	2.62	2.31	2.91	0.73
	IV	–1.00	0.13	2.66	2.40	2.99	0.71
	V	–0.55	0.41	2.56	2.15	3.25	0.74

Rules for Task Sequence and Layout

Linking tasks typically are placed in each form first; the sequence of the linking tasks is kept consistent from form to form. The initial tasks on a form and in each session are relatively easier than those tasks that follow so that many students can experience success early in each testing session. The remaining tasks are sequenced within a form and within a session by alternating easier and more difficult tasks.

Chapter 5: Test Administration

Test Security and Confidentiality

All tests within the STAR Program are secure documents. For the CAPA 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 for student data. To ensure security for all the tests that ETS develops or handles, ETS maintains an Office of Testing Integrity (OTI), which is described in the next section.

ETS's Office of Testing Integrity

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

OTI's mission is to:

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

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

Test Development

During the test development process, ETS staff members consistently adhere to the following established security procedures:

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

Task and Data Review

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

- Individuals who participate in the ARPs must sign a 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 prohibited in the meeting rooms.

Item Banking

When the ARP review is complete, the tasks are placed in the item bank. ETS then delivers the tasks to the CDE through the California electronic item bank. Subsequent updates to tasks are based on data from field testing and the operational use of the tasks. Only the latest content of the task is in the bank at any time, along with the administration data from every administration that has included the task.

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

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

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

Transfer of Forms and Tasks to the CDE

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

Security of Electronic Files Using a Firewall

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

All electronic applications included in the STAR Management System (CDE, 2011a) remain protected by the ETS firewall software at all times. Due to the sensitive nature of the student information processed by the STAR Management System, the firewall plays a significant role in maintaining an assurance of confidentiality in the users of this information. (It should

be noted that the STAR Management System neither stores nor processes tests or student test results.)

Printing and Publishing

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

Test Administration

Pearson receives testing materials from printers, packages them, and sends them to school districts. After testing, the school districts return materials to Pearson for scoring. During these events, Pearson takes extraordinary measures to protect the testing materials. Pearson's customized Oracle business applications verify that inventory controls are in place, from materials receipt to packaging. The reputable carriers used by Pearson provide a specialized handling and delivery service that maintains test security and meets the STAR 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 testing 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 STAR coordinator, who is responsible for returning them to the STAR Scoring and Processing Centers. The following measures are in place to ensure security of STAR testing materials:

- District STAR coordinators are required 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 may ship any testing materials to the school 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 may be delivered to the school/test site.
- Anyone having access to the testing materials must 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 must contact the CDE

immediately and the coordinator will be asked to follow up with a written explanation of the violation or suspected violation.

Processing and Scoring

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

- There is controlled access to the facility.
- No 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 identification badges at all times in Pearson facilities.

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

The completed and scored answer documents are stored in secure warehouses. After they are stored, they will not be handled again unless questions arise about a student's score. School and district personnel are not allowed to look at a completed answer documents unless necessary for the purpose of transcription or to investigate irregular cases.

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

Data Management

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

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

For mainframe processes, Pearson utilizes Random Access Control Facility (RACF) to limit and control access to all data files (test and production), source code, object code, databases, and tables. RACF controls who is authorized to alter, update, or even read the files. All attempts to access files on the mainframe by unauthorized users are logged and monitored. In addition, Pearson uses ChangeMan, a mainframe configuration management tool, to control versions of the software and data files. ChangeMan provides another level of security, combined with RACF, to place the correct tested version of code into production. Unapproved changes are not implemented without prior review and approval.

Transfer of Scores via Secure Data Exchange

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

Statistical Analysis

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

Reporting and Posting Results

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

- Paper reports, some with individual student results and others with summary results
- Encrypted files of summary results (sent to the CDE by means of SFTP); any summary results that have fewer than eleven students are not reported.
- Task-level statistics based on the results which are entered into the item bank

Student Confidentiality

To meet ESEA and state requirements, school districts must collect demographic data about students. This includes information about students' ethnicity, parent education, disabilities, whether the student qualifies for the National School Lunch Program (NSLP), and so forth (CDE, 2011b). ETS takes precautions to prevent any of this information from becoming public or being used for anything other than testing purposes. These procedures are applied to all documents in which these student demographic data may appear, including in Pre-ID files and reports.

Student Test Results

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

ETS has implemented network controls for routers, gateways, switches, firewalls, network tier management, and network connectivity. Routers, gateways, and switches represent points of access between networks. However, these do not contain mass storage or represent points of vulnerability, particularly to unauthorized access or denial of service. Routers, switches, firewalls, and gateways may possess little in the way of logical access.

ETS has many facilities and procedures that protect computer files. Facilities, policies, software, and procedures such as firewalls, intrusion detection, and virus control are in place to provide for physical security, data security, and disaster recovery. Comprehensive disaster recovery facilities are available and tested regularly at the SunGard installation in

Philadelphia, Pennsylvania. ETS routinely sends backup data cartridges and files for critical software, applications, and documentation to a secure off-site storage facility for safekeeping.

Access to the ETS Computer Processing Center is controlled by employee and visitor identification badges. The Center is secured by doors that can be unlocked only 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 in use at the Center.

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

- Scoring
- Transfer of scores by means of secure data exchange
- Reporting
- Posting of aggregate data
- Storage

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

Procedures to Maintain Standardization

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

Test Administrators

The CAPA is administered in conjunction with the other tests that comprise the STAR Program. ETS employs personnel who facilitate various processes involved in the standardization of an administration cycle.

The responsibilities for district and test site staff members are included in the *STAR District and Test Site Coordinator Manual* (CDE, 2011c). This manual is described in the next section.

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

District STAR Coordinator

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

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

Test Site Coordinator

At each test site, the superintendent of the school district or the district STAR coordinator designates a STAR test site coordinator from among the employees of the school district. (*5 CCR* Section 858 [a])

Test site coordinators are responsible for making sure that the school has the proper testing materials, distributing testing materials within a school, securing materials before, during, and after the administration period, answering questions from test examiners, preparing and packaging materials to be returned to the school district after testing, and returning the materials to the school district. (CDE, 2011c)

Test Examiner

The CAPA is administered to students individually by test examiners who may be assisted by test proctors and scribes. A test examiner is an employee of a school district or an employee of a nonpublic, nonsectarian school (NPS) who has been trained to administer the tests and has signed a STAR Test Security Affidavit. For the CAPA, the test examiner must be a certificated or licensed school staff member (5 *CCR* Section 850 [q]). Test examiners must follow the directions in the *CAPA Examiner's Manual* (CDE, 2011d) exactly.

Test Proctor

A test proctor is an employee of the school district or a person, assigned by an NPS to implement the IEP of a student, who has received training designed to prepare the proctor to assist the test examiner in the administration of tests within the STAR Program (5 *CCR* Section 850 [r]). Test proctors must sign STAR Test Security Affidavits (5 *CCR* Section 859 [c]).

Observer

To ensure the comparability of scores, the test site coordinator and principal of the school should objectively and randomly select ten percent of the students who will take the CAPA in each content area at each level at each site to receive a second rating. The observer is a certificated or licensed employee (5 *CCR* Section 850 [q]) who observes the administration of each task and completes a separate answer document for those students who are second-rated.

CAPA Examiner's Manual

The *CAPA Examiner's Manual* describes the CAPA administrative procedures and scoring rubrics and contains the manipulative lists and all the tasks for all the CAPA content area tests at each level. Examiners must follow task preparation guidelines exactly (CDE, 2011d).

District and Test Site Coordinator Manual

Test administration procedures are to be followed exactly so that all students have an equal opportunity to demonstrate their academic achievement. The *STAR District and Test Site Coordinator Manual* contributes to this goal by providing information about the responsibilities of district and test site coordinators, as well as those of the other staff involved in the administration cycle (CDE, 2011c). However, the manual is not intended as a

substitute for the *CCR*, *Title 5,* Education (*5 CCR*) or to detail all of the coordinator's responsibilities.

STAR Management System Manuals

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

- **Test Administration Setup—**This module allows school districts to determine and calculate dates for scheduling test administrations for school districts, to verify contact information for those school districts, and to update the school district's shipping information. (CDE, 2011e)
- **Order Management—**This module allows school districts to enter quantities of testing materials for schools. Its manual includes guidelines for determining which materials to order. (CDE, 2011f)
- **Pre-ID—**This module allows school districts to enter or upload student information including demographics and to identify the test(s) the student will take. This information is printed on student test booklets or answer documents or on labels that can be affixed to test booklets or answer documents. Its manual includes the CDE's Pre-ID layout. (CDE, 2011b)
- **Extended Data Corrections—**This module allows school districts to correct the data that were submitted during Pre-ID prior to the last day of the school district's selected testing window. (CDE, 2011g)

Accommodations for Students with Disabilities

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

Identification

Most students with disabilities and most English learners take the CSTs under standard conditions. However, some students with disabilities and some English learners may need assistance when taking the tests. This assistance takes the form of test variations, accommodations, or modifications. The Matrices of Test Variations, Accommodations, and Modifications for administrations of California Statewide Assessments are provided in Appendix E of the *STAR District and Test Site Coordinator Manual* (CDE, 2011c). Because examiners may adapt the CAPA in light of a student's instructional mode, accommodations and modifications do not apply to the CAPA.

Adaptations

Students eligible for the CAPA represent a diverse population. Without compromising the comparability of scores, adaptations are allowed on the CAPA to ensure the student's

optimal performance. These adaptations are regularly used for the student in the classroom throughout the year. The CAPA includes two types of adaptations:

1. Suggested adaptations for particular tasks, as specified in the task preparation instructions; and
2. Core adaptations, which are applicable for many of the tasks.

The core adaptations may be appropriate for students across many of the CAPA tasks and are provided in the *CAPA Examiners' Manual* (CDE, 2011d), on page 23 of the nonsecure manual.

Scoring

CAPA tasks are scored using a 5-point rubric (Level I) or a 4-point (Levels II–V) holistic rubric approved by the CDE. The rubrics include specific behavioral descriptors for each score point to minimize subjectivity in the rating process and facilitate score comparability and reliability. Student performance on each task is scored by one primary examiner, usually the child's teacher, or by another licensed or certificated staff member who is familiar to the student and who has completed the CAPA training. To establish scoring reliability, approximately ten percent of students receive a second independent rating by a trained observer who is also a licensed or certificated staff member and has completed the CAPA training. The answer document indicates whether the test was scored by the examiner or the observer.

Demographic Data Corrections

After reviewing student data, some school districts may discover demographic data or CAPA levels that are incorrect. The Demographic Data Corrections module of the STAR Management System gives school district the means to correct these data within a specified availability window. Districts may correct data to: (1) Have the school district's API/AYP recalculated; (2) Rescore uncoded or miscoded CAPA levels; (3) Obtain a corrected data CD-ROM for school district records; or (4) Match unmatched records (CDE, 2011h).

Testing Irregularities

Testing irregularities are circumstances that may compromise the reliability and validity of test results and, if more than five percent of the students tested are involved, could affect a school's API and AYP.

The district STAR coordinator is responsible for immediately notifying the CDE of any irregularities that occur before, during, or after testing. The test examiner is responsible for immediately notifying the district STAR coordinator of any security breaches or testing irregularities that occur in the administration of the test. Once the district STAR coordinator and the CDE have determined that an irregularity has occurred, the CDE instructs the district STAR coordinator on how and where to identify the irregularity on the answer document. The information and procedures to assist in identifying irregularities and notifying the CDE are provided in the *STAR District and Test Site Coordinator Manual* (CDE, 2011c).

Test Administration Incidents

A test administration incident is any event that occurs before, during, or after test administrations that does not conform to the instructions stated in the *CAPA Examiner's Manual* (CDE, 2011d) and the *STAR District and Test Site Coordinator Manual* (CDE, 2011c). These events include test administration errors and disruptions. Test administration incidents generally do not affect test results. These administration incidents are not reported

to the CDE or the STAR Program testing contractor. The STAR test site coordinator should immediately notify the district STAR coordinator of any test administration incidents that occur. It is recommended by the CDE that districts and schools maintain records of these incidents.

References

California Department of Education (2011a). *2011 STAR Management System.* http://www.startest.org/sms.html

California Department of Education (2011b). *2011 STAR Pre-ID instructions manual*. Sacramento, CA. Downloaded from http://www.startest.org/pdfs/STAR.pre-id_manual.2011.pdf

California Department of Education (2011c). *2011 STAR district and test site coordinator manual*. Sacramento, CA. Downloaded from http://www.startest.org/pdfs/STAR.coord_man.2011.pdf

California Department of Education (2011d). *2011 California Alternate Performance Assessment (CAPA) examiner's manual.* Sacramento, CA. Downloaded from http://www.startest.org/pdfs/CAPA.examiners_manual.nonsecure.2011.pdf

California Department of Education (2011e). *2011 STAR Test Administration Setup manual*. Sacramento, CA. Downloaded from http://www.startest.org/pdfs/STAR.test_admin_setup.2011.pdf

California Department of Education (2011f). *2011 STAR Order Management manual*. Sacramento, CA. Downloaded from http://www.startest.org/pdfs/STAR.order_mgmt.2011.pdf

California Department of Education (2011g). *2011 STAR Extended Data Corrections manual*. Sacramento, CA. Downloaded from http://www.startest.org/pdfs/STAR.xdc_manual.2011.pdf

California Department of Education. (2011h). *2011 STAR Demographic Data Corrections manual*. Sacramento, CA. Downloaded from http://www.startest.org/pdfs/STAR.data_corrections_manual.2011.pdf

Chapter 6: Performance Standards

Background

The CAPA was first administered in 2003. Subsequently, the CAPA has been revised to better link it to the grade-level California content standards. The revised blueprints for the CAPA were approved by the SBE in 2006 for implementation beginning in 2008; new tasks were developed to meet the revised blueprints and field tested.

From September 16 to 18, 2008, ETS conducted a standard-setting workshop in Sacramento, California, to recommend cut scores that delineated the revised performance standards for the CAPA for ELA and mathematics levels I through V and the CAPA for science levels I and III through V.(the CAPA for Science is not assessed in Level II). The performance standards were defined by the SBE as far below basic, below basic, basic, proficient, and advanced.

Performance standards are developed from a general description of the performance level (policy level descriptors) and competencies lists, which operationally define each level. Cut scores numerically define the performance levels. This chapter describes the process of developing performance standards which were first applied to the CAPA operational tests in the spring of 2009.

California employed carefully designed standard-setting procedures to facilitate the development of performance standards for each CAPA. The standard-setting method used for the CAPA was the Performance Profile Method, a holistic judgment approach based on profiles of student test performance for the areas of ELA and mathematics at all five test levels, and for science at levels I, III, IV, and V. Four panels of educators were convened to recommend cut scores: one panel for each content area focused on all levels above Level I and a separate panel focused on Level I. After the standard setting, ETS met with representatives of the CDE to review the preliminary results and provided an executive summary of the procedure and tables that showed the panel-recommended cut scores and impact data. The final cut scores were adopted by the SBE in November 2008. An overview of the standard setting workshop and final results are provided below; see the technical report for the standard setting (ETS, 2008a) for more detailed information.

Standard Setting Procedure

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

- Experience with administering the CAPA
- Direct experience in the education of students who take the CAPA
- Familiarity with the California content standards

Panelists were recruited to be representative of the educators of the state’s CAPA-eligible students (ETS, 2008b). Panelists were assigned to one of four panels (Level I, ELA, mathematics, or science) such that the educators on each panel should have experience administering CAPA across the levels in the content area(s) to which they were assigned.

As with other standard setting processes, panelists participating in the CAPA workshop followed the steps listed below.

1. Prior to attending the workshop, all panelists received a pre-workshop assignment. The task was to review, on their own, the content standards upon which the CAPA tasks are based and take notes on their own expectations for students at each performance level. This allowed the panelists to understand how their perceptions may relate to the complexity of content standards.
2. At the start of the workshop, panelists received training which included the purpose of standard setting and their role in the work, the meaning of a "cut score" and "impact data"; and specific training and practice in the method. Impact data included the percentage of students assessed in a previous test administration of the test who would fall into each performance level, given the panelists' judgments of cut scores.
3. Panelists became familiar with the tasks by reviewing the actual test and the rubrics and then assessing and discussing the demands of the tasks.
4. Panelists reviewed the draft list of competencies as a group, noting the increasing demands of each subsequent level. The competencies lists were developed by a subset of the standard-setting panelists based on the California content standards and policy level descriptors (see the next section). In this step, they began to visualize the knowledge and skills of students in each performance level and the differences between levels.
5. Panelists identified characteristics of a "borderline" test taker or "target student." This student is defined as one who possesses just enough knowledge of the content to move over the border separating a performance level from the performance level below.
6. Training in the method was completed, and panelists' understanding of the process was confirmed through an evaluation questionnaire.
7. Panelists made individual judgments and discussed feedback related to other panelists' judgments and feedback based on student performance data (impact data). Note that no impact data were presented to the Level I panel due to the change in the Level I rubric. Panelists revised their judgments during the process if they wished.
8. The final recommended cut scores were based on an average of panelists' judgments at the end of three rounds. For the CAPA, the cut scores recommended by the panelists and the recommendation of the State Superintendent of Public Instruction were presented for public comment at regional public hearings. Comments and recommendations were then presented to the SBE for approval.

Development of Competencies Lists

Prior to the CAPA standard-setting workshop, ETS facilitated a meeting in which a subset of the standard-setting panelists was assembled to develop a list of competencies based on the California content standards and policy level descriptors. Four panels of educators were assembled to identify and discuss the competencies required of students in the CAPA levels and content areas for each performance level (below basic, basic, proficient, and advanced). Panels consisted of educators with experience working with students who take the CAPA. Panelists were assigned to one of four panels (Level I, ELA, mathematics, or science) based on experience working with students and administering the CAPA. At the conclusion of the meeting, the CDE reviewed the draft lists and delivered the final lists for

use in standard setting. The lists were used to facilitate the discussion and construction of the target student definitions during the standard-setting workshop.

Standard Setting Methodology

Performance Profile Method

Because of the small number of tasks and the fact that all CAPA tasks are constructed response items, ETS applied a procedure that combined the Policy Capturing Method (Plake & Hambleton, 2001; Jaeger, 1995a; Jaeger, 1995b) and the Dominant Profile Method (Plake & Hambleton, 2001; Plake, Hambleton, & Jaeger, 1997; Putnam, Pence, & Jaeger, 1995). Both methods are holistic methods in that they ask panelists to make decisions based on an examinee's score profile or performance rather than on each separate item.

The combined procedure that was used in 2008 is called the Performance Profile Method in this report. The procedure was a modification to the Performance Profile Method used for the CAPA standard setting in 2003 (CDE, 2003). The task for panelists was to mark the raw score representing the competencies a student should have at each performance level, that is, basic, proficient, and advanced; cut scores for below basic and far below basic performance levels were set statistically.

For each test, materials were developed so that panelists could review score patterns, or performance profiles, for the eight CAPA tasks; panelists used the profiles and corresponding raw scores to make cut score judgments. Profiles for Levels II–V were selected using 2008 student performance data. Profiles for Level I were informed by 2008 student performance data; however, due to a change in the Level I rubric after the 2008 test administration, the selection of Level I profiles also relied on verification by CAPA assessment experts, taking into account the changes in the Level I rubric (see Chapter 7 for more information on the rubric change).

The student profiles were presented at selected raw score points in an increasing order. For most raw score points, two to three profiles are presented; but in the portion of the score range where total scores are achieved by a large group of students as indicated by the operational data, up to five profiles are presented. While it is recognized that any number of combinations of item ratings may result in the same total raw scores, the intent in the Performance Profile Method is to use a cut score that is compensatory in nature. Therefore, profiles within the same total raw score are ordered randomly. Panelists are instructed that it is permissible to select total raw scores "between" the presented raw score profiles as their recommended cut score judgment for any level.

More details regarding the process implemented for the CAPA standard setting and results summary can be found in the standard-setting technical report (ETS, 2008a).

Results

The cut scores obtained as a result of the standard setting process were expressed in terms of raw scores; the panel median score after three rounds of judgments is the cut score recommendation for each level. These scores were transformed to scale scores that range between 15 and 60.

The cut score for the basic performance level was set equal to a scale score of 30 for every test level and content area; this means that a student must earn a score of 30 or higher to achieve a basic classification. The cut score for the proficient level was set equal to 35 for

each test level and content area; this means that a student must earn a score of 35 or higher to achieve a proficient classification.

The cut scores for the other performance levels usually vary by grade and content area. They are derived using procedures based on item response theory (IRT). Please note that in the case of polytomously scored items, the IRT test characteristic function is the sum of the item response functions (IRF), where the IRF of an item is the weighted sum of the response functions for each score category (weighted by the scores of the categories).

Each raw cut score for a given test is mapped to an IRT *theta* (θ) using the test characteristic function and then transformed to the scale score metric using the following equation:

$$\text{Scale Cut Score} = (35 - \theta_{proficient} \times \left(\frac{35-30}{\theta_{proficient} - \theta_{basic}}\right)) + \left(\frac{35-30}{\theta_{proficient} - \theta_{basic}}\right) \times \theta \qquad (6.1)$$

where,

θ represents the student ability,

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

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

The scale-score ranges for each performance level are presented in Table 2.2 on page 14. The cut score for each performance level is the lower bound of each scale score range. The scale score ranges do not change from year to year. Once established, they remain unchanged from administration to administration until such time that new performance standards are adopted.

Table 7.5 on page 51 in Chapter 7 presents the percentages of examinees meeting each performance level in 2011.

References

Educational Testing Service. (2003). *CAPA standard setting technical report* (California Department of Education Contract Number 5417). Princeton, NJ: Author.

Educational Testing Service. (2008a). *Technical report on the standard setting workshop for the California Alternate Performance Assessment, December 29, 2008,* (California Department of Education Contract Number 5417). Princeton, NJ: Author.

Educational Testing Service K–12 Statistical Analysis Group. (2008b). *A study to examine the effects of changes to the CAPA Level I rubric involving the hand-over-hand prompt*, Unpublished memorandum. Princeton, NJ: Author.

Jaeger, R.M. (1995a). Setting performance standards through two-stage judgmental policy capturing. *Applied Measurement in Education,* 8, pp. 15–40.

Jaeger, R.M. (1995b). Setting standards for complex performances: An iterative, judgmental policy-capturing strategy. *Educational Measurement: Issues and Practice,* 14 (4), pp. 16–20.

Plake, B. S., & Hamilton, R.K. (2001). The analytic judgment method for setting standards on complex performance assessments. In G.J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives,* (pp. 283–312). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Plake, B., Hamilton, R., & Jaeger, R.M. (1997). A new standard setting method for performance assessments: The dominant profile judgment method and some field-test results. *Educational and Psychological Measurement,* 57, pp. 400–11.

Putnam, S.E., Pence, P. & Jaeger, R.M. (1995). A multi-stage dominant profile method for setting standards on complex performance assessments. *Applied Measurement in Education,* 8, pp. 57–83.

Chapter 7: Scoring and Reporting

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

Procedures for Maintaining and Retrieving Individual Scores

The CAPA is composed entirely of performance tasks. Each content area includes eight performance tasks that are scored by a trained examiner using a rubric that depends on the test level being assessed. After the student has responded to a task, the examiner marks the score using the corresponding circle on the student's answer document.

Scoring Rubric

The scoring rubric represents the guideline for scoring the task. The rubric varies according to the CAPA level. The rubric for CAPA Level I has a range of 0–5, with 5 being the maximum score. The rubric for CAPA Levels II–V has a range of 0–4, with 4 being the maximum score.

Beginning with the administration of the 2009 CAPA, the Level I rubric was changed to take into account issues related to scoring students who required a hand-over-hand prompt (ETS, 2008). ETS believed there was a significant difference between levels of prompting when dealing with this special population of students as evidenced by the amount of special education research that deals exclusively with prompting hierarchies. A child with significant cognitive disabilities who is able to complete a task successfully at one level of prompting may take weeks or months to increase his or her proficiency in that task in order to be able to complete the task successfully at a less intrusive level of prompting. The differences within prompting levels are the reason why ETS supported a rubric that differentiates between levels of prompting and scores the responses accordingly. For Level I ELA, mathematics, and science, all tasks are scored using the same rubric. For all other levels, the rubric is specific to the task. Both rubrics are presented in Table 7.1. Note that a score of zero in Level I indicates that the student did not orient toward a task after multiple prompts had been utilized. In Levels II–V, a score of zero implies that the student did not attempt the task. In both cases, the score is defined as "No Response" for the purpose of scoring the task.

Table 7.1 Rubrics for CAPA Scoring

Level I		Levels II–V	
Score Points	**Description**	**Score Points**	**Description**
5	Correct with no prompting		
4	Correct with verbal or gestural prompt	4	Completes task with 100 percent accuracy
3	Correct with modeled prompt	3	Partially completes task (as defined for each task)
2	Correct with hand-over-hand prompt (student completes task independently)	2	Minimally completes task (as defined for each task)

Level I		Levels II–V	
Score Points	**Description**	**Score Points**	**Description**
1	Orients to task or incorrect response after attempting the task independently	1	Attempts task
0	No response	0	Does not attempt task

In order to score and report CAPA results, ETS follows an established set of written procedures. These specifications are presented in the next sections.

Scoring and Reporting Specifications

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

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

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

Scanning and Scoring

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

Each school district must return scorable and nonscorable materials within five working days after the selected last day of testing for each test administration period.

Types of Scores

Raw Score

For the CAPA for ELA and mathematics, there are five test levels and eight tasks per level. For the CAPA for science, there are four test levels and eight tasks per level. Performance scoring for Level I is based on a rubric with a range of 0–5 with a maximum score of 5. Performance scoring for Levels II–V is based on a rubric with a range of 0–4 with a maximum score of 4. The raw scores for Level I range from 0 to 40; for the other CAPA levels, the raw scores range is from 0 to 32.

Scale Score

Raw scores obtained on each CAPA test are converted to two-digit scale scores using the calibration process described in Chapter 2 on page 12. Scale scores range from 15 to 60 on each CAPA content-area test. The scale scores of examinees that have been tested in different years at a given CAPA test level and content area can be compared. However, the raw scores of these examinees cannot be meaningfully compared, because these scores are affected by the relative difficulty of the test taken as well as the ability of the examinee.

Performance Levels

For the CAPA content-area tests, the performance of each student is categorized into one of the following performance levels:

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

For all CAPA tests, the cut score for the basic performance level is 30; this means that a student must earn a scale score of 30 or higher to achieve a basic classification. The cut score for the proficient performance level is 35; this means that a student must earn a scale score of 35 or higher to achieve a proficient classification. The cut scores for the other performance levels usually vary by level and content area.

Score Verification Procedures

Various necessary measures are taken to ascertain that the student scores are computed accurately.

Monitoring and Quality Control of Scoring

Scorer Selection

Careful consideration is given to the selection of examiners for proper administration and scoring of the CAPA. It is preferred that the special education teacher or case carrier who regularly works with the student being tested administer and score the test. The examiner is required to be certificated or licensed and have successfully completed comprehensive training on CAPA administration.

If the examiner or case carrier is not available to administer the test, it may be administered and scored by another CAPA-trained staff member such as a school psychologist; speech, physical, or occupational therapist; program specialist; or certified teacher, principal or assistant principal. This individual should have experience working with students with significant cognitive disabilities and must be trained to administer the CAPA (CDE, 2011a).

Quality Control

Each student's responses to the CAPA tasks are rated by a single examiner; the total score is based on that rater's ratings. In addition, approximately 10 percent of students at each test site are also rated by an observer to provide data that can be used to assess the accuracy and reliability of the scores. The observer, who is expected to meet the same qualification requirements as an examiner, scores the test at the same time as the test is being administered, but independently of the examiner. The score from the observer does not count toward the student's CAPA score.

Score Verification Process

ETS psychometricians employ special procedures that adjust for differences in item difficulty of one test form to another. (See Chapter 2, Equating, for details.) As a result of this process, scoring tables are produced. Such tables map the current year's raw score to an appropriate scale score. A series of quality control (QC) checks are carried out by ETS psychometricians to ensure the accuracy of each scoring table, as discussed in Chapter 9 on page 166.

Pearson utilizes the scoring tables to generate scale scores for each student. ETS verifies Pearson's scale scores by conducting QC and reasonableness checks, which are described in Chapter 9 on page 167.

Overview of Score Aggregation Procedures

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

Individual Scores

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

Score Distributions and Summary Statistics

Summary statistics that describes student performance on each CAPA are presented in Table 7.2 through Table 7.4. Included in these tables are the number of tasks in each test, the number of examinees taking each test, and the means and standard deviations of student scores expressed in terms of both raw scores and scale scores. In addition, summary statistics for the operational tasks on each test are provided.

Table 7.2 Summary Statistics Describing Student Scores: ELA

Level	I	II	III	IV	V
		Scale Score Information			
Number of examinees	13,719	6,643	7,112	9,858	10,217
Mean score	40.79	38.72	39.53	39.20	38.88
SD *	10.63	5.70	6.17	7.44	6.44
Possible range	15–60	15–60	15–60	15–60	15–60
Obtained range	15–60	15–60	15–60	15–60	15–60
Median	41	39	39	40	39
Reliability	0.88	0.82	0.88	0.86	0.89
SEM †	3.68	2.40	2.17	2.78	2.09

Level	I	II	III	IV	V
Raw Score Information					
Mean score	25.69	20.00	22.78	20.24	21.84
SD *	11.16	6.14	6.24	6.65	6.73
Possible range	0–40	0–32	0–32	0–32	0–32
Obtained range	0–40	0–32	0–32	0–32	0–32
Median	28	20	24	21	23
Reliability	0.88	0.82	0.88	0.86	0.89
SEM †	3.86	2.59	2.19	2.48	2.19
Task Information					
Number of tasks	8	8	8	8	8
Mean AIS ‡	3.21	2.51	2.86	2.54	2.74
SD AIS ‡	0.51	0.50	0.55	0.53	0.49
Min. AIS	2.73	1.91	2.21	1.64	1.97
Max. AIS	4.24	3.57	3.63	3.49	3.50
Possible range	0–5	0–4	0–4	0–4	0–4
Mean polyserial	0.79	0.72	0.78	0.77	0.80
SD polyserial	0.04	0.06	0.06	0.04	0.08
Min. polyserial	0.71	0.60	0.68	0.71	0.67
Max. polyserial	0.84	0.80	0.85	0.84	0.89
Mean Rasch difficulty	–0.61	–0.99	–1.32	–1.05	–1.16
SD Rasch difficulty	0.27	0.60	0.89	0.64	0.68
Min. Rasch difficulty	–1.18	–2.18	–2.77	–2.30	–2.19
Max. Rasch difficulty	–0.40	–0.28	–0.30	0.01	–0.22

* Standard Deviation | † Standard Error of Measurement | ‡ Average Item (Task) Score

Table 7.3 Summary Statistics Describing Student Scores: Mathematics

Level	I	II	III	IV	V
Scale Score Information					
Number of examinees	13,689	6,624	7,098	9,845	10,196
Mean score	36.08	37.45	36.33	37.26	37.32
SD *	8.70	7.64	4.99	8.11	7.63
Possible range	15–60	15–60	15–60	15–60	15–60
Obtained range	15–60	15–60	15–60	15–60	15–60
Median	37	37	37	37	37
Reliability	0.84	0.82	0.77	0.85	0.85
SEM †	3.43	3.22	2.41	3.11	2.96
Raw Score Information					
Mean score	22.79	20.34	18.57	19.65	20.04
SD *	10.82	6.28	5.46	6.89	7.04
Possible range	0–40	0–32	0–32	0–32	0–32
Obtained range	0–40	0–32	0–32	0–32	0–32
Median	24	20	19	20	20
Reliability	0.84	0.82	0.77	0.85	0.85
SEM †	4.27	2.64	2.63	2.64	2.73

Level	I	II	III	IV	V
Task Information					
Number of tasks	8	8	8	8	8
Mean AIS ‡	2.85	2.55	2.32	2.47	2.51
SD AIS ‡	0.17	0.67	0.39	0.49	0.41
Min. AIS	2.66	1.38	1.76	1.83	2.01
Max. AIS	3.23	3.49	2.98	3.00	3.35
Possible range	0–5	0–4	0–4	0–4	0–4
Mean polyserial	0.75	0.73	0.68	0.75	0.76
SD polyserial	0.02	0.09	0.07	0.10	0.03
Min. polyserial	0.72	0.59	0.62	0.61	0.72
Max. polyserial	0.78	0.84	0.80	0.85	0.81
Mean Rasch difficulty	–0.21	–1.02	–0.87	–0.86	–0.98
SD Rasch difficulty	0.12	0.68	0.30	0.48	0.34
Min. Rasch difficulty	–0.43	–1.91	–1.33	–1.47	–1.54
Max. Rasch difficulty	0.00	0.18	–0.50	–0.05	–0.43

* Standard Deviation | † Standard Error of Measurement | ‡ Average Item (Task) Score

Table 7.4 Summary Statistics Describing Student Scores: Science

Level	I	III	IV	V
Scale Score Information				
Number of examinees	3,512	3,391	3,155	3,245
Mean score	36.20	36.42	36.23	35.82
SD *	10.67	5.27	5.72	5.08
Possible range	15–60	15–60	15–60	15–60
Obtained range	15–60	15–60	15–60	15–60
Median	36	36	36	36
Reliability	0.89	0.84	0.84	0.86
SEM †	3.55	2.09	2.28	1.92
Raw Score Information				
Mean score	23.23	21.42	21.40	20.59
SD *	11.72	6.26	5.82	5.99
Possible range	0–40	0–32	0–32	0–32
Obtained range	0–40	0–32	0–32	0–32
Median	24	22	22	21
Reliability	0.89	0.84	0.84	0.86
SEM †	3.90	2.49	2.32	2.26
Task Information				
Number of tasks	8	8	8	8
Mean AIS ‡	2.91	2.69	2.69	2.58
SD AIS ‡	0.27	0.20	0.25	0.37
Min. AIS	2.65	2.34	2.34	2.14
Max. AIS	3.50	2.94	2.96	3.33
Possible range	0–5	0–4	0–4	0–4
Mean polyserial	0.80	0.73	0.73	0.76
SD polyserial	0.02	0.06	0.03	0.04

Level	I	III	IV	V
Min. polyserial	0.76	0.61	0.67	0.69
Max. polyserial	0.83	0.80	0.75	0.81
Mean Rasch difficulty	–0.33	–1.29	–1.19	–0.52
SD Rasch difficulty	0.14	0.33	0.30	0.40
Min. Rasch difficulty	–0.64	–1.90	–1.56	–1.32
Max. Rasch difficulty	–0.18	–0.88	–0.80	–0.09

* Standard Deviation | † Standard Error of Measurement | ‡ Average Item (Task) Score

The percentages of students in each performance level are presented in Table 7.5. The numbers in the summary tables may not match exactly the results reported on the CDE Web site because of slight differences in the samples used to compute the statistics. The P2 data file was used for the analyses in this chapter. This file contained data collected from all school districts but did not include corrections of demographic data through the Demographic Data Corrections process. In addition, students with invalid scores were excluded from the tables.

Table 7.5 Percentage of Examinees in Each Performance Level

Content Area	CAPA Level	Far Below Basic	Below Basic	Basic	Proficient	Advanced
English–Language Arts	I	5%	6%	9%	24%	56%
	II	1%	4%	10%	43%	42%
	III	1%	3%	12%	38%	46%
	IV	1%	6%	14%	44%	35%
	V	1%	4%	17%	31%	48%
Mathematics	I	7%	8%	16%	36%	33%
	II	2%	9%	23%	31%	35%
	III	1%	5%	21%	54%	18%
	IV	2%	13%	18%	35%	32%
	V	2%	8%	22%	34%	34%
Science	I	7%	12%	22%	23%	35%
	III	1%	4%	26%	50%	18%
	IV	1%	6%	29%	45%	20%
	V	1%	5%	25%	48%	20%

Table 7.A.1 through Table 7.A.3 in Appendix 7.A, starting on page 57, show the distributions of scale scores for each CAPA. The results are reported in terms of three score intervals. A cell value of “N/A” indicates that there are no obtainable scale scores within that scale score range for the particular CAPA.

Group Scores

Statistics summarizing student performance by content area for selected groups of students are provided in Table 7.B.1 through Table 7.B.3 for the CAPA.

In the tables, students are grouped by demographic characteristics, including gender, ethnicity, English-language fluency, economic status, and primary disability. The tables show, for each demographic group, the numbers of valid cases and percentages of students in each performance level by demographic group.

Table 7.6 provides definitions of the demographic groups included in the tables. Students’ economic status was determined by considering the education level of their parents and whether or not they participated in the National School Lunch Program (NSLP). To protect

privacy when the number of students in a subgroup is 10 or fewer, the summary statistics at the test level are not reported and are presented as hyphens.

Table 7.6 Subgroup Definitions

Subgroup	Definition
Gender	• Male • Female
Ethnicity	• African American • American Indian or Alaska Native • Asian – Asian Indian – Cambodian – Chinese – Hmong – Japanese – Korean – Laotian – Vietnamese – Other Asian • Hispanic or Latino • Pacific Islander – Guamanian – Native Hawaiian – Samoan – Tahitian – Other Pacific Islander • Filipino • White (not Hispanic)
English-language Fluency	• English only • Initially fluent English proficient • English learner • Reclassified fluent English proficient
Economic Status	• Not economically disadvantaged • Economically disadvantaged
Primary Disability	• Mental retardation • Hard of hearing • Deafness • Speech/language impairment • Visual impairment • Emotional disturbance • Orthopedic impairment • Other health impairment • Specific learning impairment • Deaf blindness • Multiple group • Autism • Traumatic brain injury

Reports Produced and Scores for Each Report

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

1. Communicating with parents and guardians;
2. Informing decisions needed to support student achievement;
3. Evaluating school programs; and
4. Providing data for state and federal accountability programs for schools and districts.

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

Types of Score Reports

There are three categories of CAPA reports. These categories and the specific reports in each category are given in the Table 7.7.

Table 7.7 Types of CAPA Reports

1. Summary Reports	▪ STAR Student Master List Summary ▪ STAR Subgroup Summary (including the Ethnicity for Economic Status)
2. Individual Reports	▪ STAR Student Record Label ▪ STAR Student Master List ▪ STAR Student Report for the CAPA
3. Internet Reports	▪ CAPA Scores (state, county, district, school) ▪ CAPA Summary Scores (state, county, district, school)

These reports are sent to the independently testing charter schools, counties, or school districts; the school district forwards the appropriate reports to test sites or, in the case of the STAR Student Report, sends the report(s) to the child's parent or guardian and forwards a copy to the student's school or test site. Reports such as the STAR Student Report, Student Record Label, and Student Master List that include individual student results are not distributed beyond the student's school. Internet reports are described on the CDE Web site and are accessible to the public online at http://star.cde.ca.gov/.

Score Report Contents

The STAR Student Report provides scale scores and performance levels for each CAPA taken by the student. Scale scores are reported on a scale ranging from 15 to 60. The performance levels reported are: far below basic, below basic, basic, proficient, and advanced.

Further information about the STAR Student Report and the other reports is provided in Appendix 7.C.

Score Report Applications

CAPA results provide parents and guardians with information about their children's progress. The results are a tool for increasing communication and collaboration between parents or guardians and teachers. Along with report cards from teachers and information from school and classroom tests, the STAR 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.

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

CAPA results help school districts and schools identify strengths and weaknesses in their instructional programs. Each year, school districts and school staff examine CAPA results at each level and content area tested. Their findings are used to help determine:

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

The results from the CAPA are used for state and federal accountability programs to monitor each school's and district's progress toward achieving established goals. As mentioned previously, CAPA results are used to calculate each school's and district's Academic Performance Index (API). The API is a major component of California's Public School Accountability Act (PSAA) and is used to rank the academic performance of schools, compare schools with similar characteristics (for example, size and ethnic makeup), identify low-performing and high-priority schools, and set yearly targets for academic growth.

CAPA results also are used to comply with federal ESEA legislation that requires all schools to meet specific academic goals. The progress of each school toward achieving these goals is provided annually in an Adequate Yearly Progress (AYP) report. Each year, California schools and districts must meet AYP goals by showing that a specified percentage of CAPA test-takers at the district and school level are performing at or above the proficient level on the CAPA for ELA and mathematics.

Criteria for Interpreting Test Scores

A school district may use CAPA results to help make decisions about student placement, promotion, retention, or other considerations related to student achievement. However, it is important to remember that a single test can provide only limited information. Other relevant information should be considered as well. It is advisable for parents to evaluate their child's strengths and weaknesses in the relevant topics by reviewing classroom work and progress reports in addition to the child's CAPA results (CDE, 2011b). It is also important to note that a student's score in a content area contains measurement error and could vary somewhat if the student were retested.

Criteria for Interpreting Score Reports

The information presented in various reports must be interpreted with caution when making performance comparisons. When comparing scale score and performance-level results for the CAPA, the user is limited to comparisons within the same content area and level. This is because the score scales are different for each content area and level. The user may compare scale scores for the same content area and level, within a school, between schools, or between a school and its district, its county, or the state. The user can also make comparisons within the same level and content area across years. Comparing scores obtained in different levels or content areas should be avoided because the results are not on the same scale. Comparisons between raw scores should be limited to comparisons within not only content area and level but also test year. Since new score scales and cut scores were applied beginning with the 2009 test results, results from this and subsequent

years cannot meaningfully be compared to results obtained in prior years. For more details on the criteria for interpreting information provided on the score reports, see the *2011 STAR Post-Test Guide* (CDE, 2011c).

References

California Department of Education. (2011a), *2011 CAPA examiner's manual.* Downloaded from http://www.startest.org/pdfs/CAPA.examiners_manual.nonsecure.2011.pdf

California Department of Education. (2011b). *2011 STAR CST/CMA, CAPA, and STS printed reports*. Downloaded from http://www.startest.org/pdfs/STAR.reports.2011.pdf

California Department of Education. (2011c). *2011 STAR post-test guide.* Downloaded from http://www.startest.org/pdfs/STAR.post-test_guide.2011.pdf

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

Educational Testing Service. (2008) *A study to examine the effects of changes to the CAPA Level I rubric involving the hand-over-hand prompt,* Unpublished memorandum, Princeton, NJ: Author.

Appendix 7.A—Scale Score Distribution Tables

Table 7.A.1 Scale Score Frequency Distributions: ELA, Levels I–V

Scale Score	ELA I		ELA II		ELA III		ELA IV		ELA V	
	Freq	Pct.	Freq	Pct.	Freq	Pct.	Freq	Pct.	Freq	Pct.
60	1,534	11.18	70	1.05	158	2.22	213	2.16	266	2.60
57–59	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
54–56	488	3.56	N/A	N/A	N/A	N/A	235	2.38	N/A	N/A
51–53	N/A	N/A	91	1.37	N/A	N/A	N/A	N/A	N/A	N/A
48–50	580	4.23	156	2.35	300	4.22	664	6.74	346	3.39
45–47	1,597	11.64	464	6.98	862	12.12	865	8.77	1,029	10.07
42–44	1,768	12.89	978	14.72	1,004	14.12	1,475	14.96	1,320	12.92
39–41	2,560	18.66	1,803	27.14	1,809	25.44	2,087	21.17	2,493	24.40
36–38	1,883	13.73	1,487	22.38	1,360	19.12	1,795	18.21	2,190	21.43
33–35	953	6.95	1,060	15.96	1,030	14.48	1,183	12.00	1,284	12.57
30–32	853	6.22	230	3.46	326	4.58	611	6.20	782	7.65
27–29	428	3.12	129	1.94	115	1.62	340	3.45	250	2.45
24–26	151	1.10	64	0.96	79	1.11	109	1.11	91	0.89
21–23	123	0.90	28	0.42	21	0.30	101	1.02	50	0.49
18–20	156	1.14	24	0.36	16	0.22	37	0.38	32	0.31
15–17	645	4.70	59	0.89	32	0.45	143	1.45	84	0.82

Table 7.A.2 Scale Score Frequency Distributions: Mathematics, Levels I–V

Scale Score	Math I		Math II		Math III		Math IV		Math V	
	Freq	Pct.	Freq	Pct.	Freq	Pct.	Freq	Pct.	Freq	Pct.
60	534	3.90	92	1.39	31	0.44	202	2.05	404	3.96
57–59	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
54–56	N/A	N/A	90	1.36	N/A	N/A	256	2.60	N/A	N/A
51–53	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
48–50	245	1.79	143	2.16	47	0.66	296	3.01	269	2.64
45–47	247	1.80	613	9.25	146	2.06	656	6.66	316	3.10
42–44	1,235	9.02	1,071	16.17	497	7.00	1,246	12.66	1,158	11.36
39–41	2,286	16.70	941	14.21	1,384	19.50	1,585	16.10	1,849	18.13
36–38	3,636	26.56	1,052	15.88	2,550	35.93	1,957	19.88	2,439	23.92
33–35	2,525	18.45	1,109	16.74	1,582	22.29	1,087	11.04	1,745	17.11
30–32	973	7.11	803	12.12	413	5.82	1,102	11.19	1,022	10.02
27–29	609	4.45	317	4.79	198	2.79	777	7.89	286	2.81
24–26	174	1.27	97	1.46	81	1.14	219	2.22	434	4.26
21–23	159	1.16	142	2.14	58	0.82	190	1.93	48	0.47
18–20	161	1.18	29	0.44	18	0.25	54	0.55	37	0.36
15–17	905	6.61	125	1.89	93	1.31	218	2.21	189	1.85

Table 7.A.3 Scale Score Frequency Distributions: Science, Levels I–V

Scale Score	Science I		Science III		Science IV		Science V	
	Freq	Pct.	Freq	Pct.	Freq	Pct.	Freq	Pct.
60	293	8.34	69	2.03	61	1.93	33	1.02
57–59	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
54–56	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
51–53	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
48–50	90	2.56	N/A	N/A	N/A	N/A	N/A	N/A
45–47	73	2.08	N/A	N/A	71	2.25	46	1.42
42–44	338	9.62	228	6.72	225	7.13	167	5.15
39–41	441	12.56	531	15.66	431	13.66	413	12.73
36–38	706	20.10	1,154	34.03	858	27.19	1,121	34.55
33–35	594	16.91	941	27.75	916	29.03	903	27.83
30–32	299	8.51	300	8.85	382	12.11	353	10.88
27–29	164	4.67	92	2.71	106	3.36	133	4.10
24–26	125	3.56	34	1.00	53	1.68	23	0.71
21–23	49	1.40	16	0.47	15	0.48	16	0.49
18–20	43	1.22	11	0.32	13	0.41	4	0.12
15–17	297	8.46	15	0.44	24	0.76	33	1.02

Appendix 7.B—Demographic Summaries

Table 7.B.1 Demographic Summary for ELA, All Examinees

		Percentage in Performance Level				
	Number Tested	Far Below Basic	Below Basic	Basic	Proficient	Advanced
All valid scores	47,549	2%	5%	12%	34%	46%
Male	31,022	2%	5%	12%	35%	46%
Female	16,361	2%	5%	12%	34%	47%
Gender unknown	166	2%	5%	11%	36%	45%
American Indian	404	2%	3%	9%	30%	55%
Asian American	3,141	3%	6%	16%	38%	37%
Pacific Islander	251	2%	7%	10%	32%	48%
Filipino	1,408	2%	7%	15%	37%	38%
Hispanic	24,550	2%	4%	12%	34%	47%
African American	4,685	2%	4%	10%	33%	51%
White	11,592	2%	5%	12%	34%	47%
Ethnicity unknown	1,518	2%	5%	12%	37%	44%
English Only	29,288	2%	5%	12%	34%	46%
Initially–Fluent English Proficient	805	3%	7%	17%	34%	39%
English Learner	15,331	2%	5%	12%	35%	47%
Reclassified–Fluent English Proficient	1,485	1%	4%	14%	34%	47%
English Proficient unknown	640	2%	4%	12%	33%	49%
Mental Retardation	19,341	1%	4%	13%	36%	45%
Hard of Hearing	298	2%	3%	12%	37%	46%
Deafness	357	1%	3%	17%	44%	35%
Speech/Language Impairment	1,715	0%	0%	3%	31%	66%
Visual Impairment	497	7%	9%	11%	31%	42%
Emotional Disturbance	303	1%	1%	3%	22%	73%
Orthopedic Impairment	4,040	6%	9%	14%	34%	38%
Other Health Impairment	2,06	2%	3%	8%	31%	57%
Specific Learning Impairment	2,865	0%	0%	1%	20%	79%
Deaf Blindness	40	10%	15%	23%	33%	20%
Multiple Group	2,235	8%	9%	16%	33%	34%
Autism	12,843	2%	6%	14%	37%	41%
Traumatic Brain Injury	301	4%	2%	7%	31%	55%
Unknown	708	2%	5%	10%	36%	47%
Not Econ. Disadvantaged	16,038	3%	6%	14%	35%	42%
Economically Disadvantaged	30,293	2%	4%	11%	34%	49%
Unknown Economic Status	1,218	2%	5%	10%	31%	53%
Primary Ethnicity—Not Economically Disadvantaged						
American Indian	137	4%	7%	12%	31%	45%
Asian American	1,676	3%	7%	17%	38%	36%
Pacific Islander	92	3%	9%	10%	29%	49%
Filipino	873	3%	8%	15%	39%	35%
Hispanic	4,512	4%	7%	13%	33%	43%
African American	1,294	3%	6%	12%	33%	45%
White	6,830	2%	5%	14%	36%	43%
Ethnicity unknown	624	2%	5%	13%	38%	42%

		Percentage in Performance Level				
	Number Tested	Far Below Basic	Below Basic	Basic	Proficient	Advanced
Primary Ethnicity—Economically Disadvantaged						
American Indian	260	1%	2%	7%	30%	60%
Asian American	1,385	3%	6%	15%	38%	39%
Pacific Islander	154	2%	6%	11%	34%	46%
Filipino	493	2%	5%	15%	34%	43%
Hispanic	19,589	2%	4%	12%	34%	48%
African American	3,274	2%	3%	10%	33%	52%
White	4,452	2%	4%	9%	32%	53%
Ethnicity unknown	686	2%	6%	12%	37%	43%
Primary Ethnicity—Unknown Economic Status						
American Indian	NA	NA	NA	NA	NA	NA
Asian American	80	1%	10%	18%	33%	39%
Pacific Islander	NA	NA	NA	NA	NA	NA
Filipino	42	0%	10%	14%	29%	48%
Hispanic	449	2%	4%	10%	30%	54%
African American	117	1%	5%	7%	33%	54%
White	310	3%	2%	9%	30%	56%
Ethnicity unknown	208	3%	6%	7%	35%	50%

* Results for groups with fewer than 11 members are not reported.

Table 7.B.2 Demographic Summary for Mathematics, All Examinees

		Percentage in Performance Level				
	Number Tested	Far Below Basic	Below Basic	Basic	Proficient	Advanced
All valid scores	47,452	3%	9%	20%	37%	31%
Male	30,950	3%	9%	19%	37%	32%
Female	16,336	3%	9%	21%	38%	29%
Gender unknown	166	4%	7%	20%	34%	34%
American Indian	404	1%	7%	18%	37%	37%
Asian American	3,129	4%	10%	22%	36%	29%
Pacific Islander	251	4%	8%	19%	38%	31%
Filipino	1,404	4%	11%	22%	37%	27%
Hispanic	24,505	3%	9%	19%	37%	32%
African American	4,669	3%	8%	18%	38%	33%
White	11,574	3%	9%	21%	38%	29%
Ethnicity unknown	1,516	3%	9%	21%	38%	29%
English Only	29,222	3%	9%	20%	38%	30%
Initially–Fluent English Proficient	803	4%	12%	23%	34%	27%
English Learner	15,305	3%	8%	18%	37%	33%
Reclassified–Fluent English Proficient	1,482	2%	11%	18%	34%	36%
English Proficient unknown	640	2%	7%	20%	38%	33%
Mental Retardation	19,310	2%	10%	22%	38%	28%
Hard of Hearing	298	3%	5%	17%	37%	38%
Deafness	357	1%	6%	12%	37%	44%
Speech/Language Impairment	1,711	0%	1%	10%	41%	48%
Visual Impairment	498	9%	12%	22%	33%	24%
Emotional Disturbance	299	1%	2%	12%	26%	59%
Orthopedic Impairment	4,031	9%	13%	22%	33%	23%
Other Health Impairment	2,002	2%	6%	16%	39%	37%
Specific Learning Impairment	2,865	0%	1%	5%	34%	60%
Deaf Blindness	39	10%	33%	21%	21%	15%
Multiple Group	2,229	11%	15%	22%	34%	19%
Autism	12,812	3%	8%	20%	39%	30%
Traumatic Brain Injury	299	6%	5%	19%	29%	40%
Unknown	702	2%	7%	21%	39%	31%
Not Econ. Disadvantaged	16,010	4%	10%	21%	38%	27%
Economically Disadvantaged	30,226	3%	8%	19%	37%	33%
Unknown Economic Status	1,216	3%	7%	18%	38%	34%
Primary Ethnicity—Not Economically Disadvantaged						
American Indian	137	3%	9%	21%	35%	32%
Asian American	1,669	4%	10%	22%	36%	28%
Pacific Islander	92	7%	8%	24%	32%	30%
Filipino	869	3%	12%	20%	38%	26%
Hispanic	4,507	6%	10%	21%	37%	27%
African American	1,289	5%	11%	16%	40%	27%
White	6,825	4%	10%	23%	38%	26%
Ethnicity unknown	622	4%	8%	22%	39%	27%

		Percentage in Performance Level				
	Number Tested	Far Below Basic	Below Basic	Basic	Proficient	Advanced
Primary Ethnicity—Economically Disadvantaged						
American Indian	260	0%	7%	17%	38%	38%
Asian American	1,381	4%	9%	21%	37%	29%
Pacific Islander	154	3%	8%	16%	42%	31%
Filipino	493	4%	9%	24%	35%	27%
Hispanic	19,549	2%	8%	19%	37%	33%
African American	3,264	3%	6%	18%	37%	36%
White	4,440	3%	8%	19%	37%	34%
Ethnicity unknown	685	3%	9%	20%	38%	30%
Primary Ethnicity—Unknown Economic Status						
American Indian	0	–	–	–	–	–
Asian American	79	3%	8%	24%	34%	32%
Pacific Islander	0	–	–	–	–	–
Filipino	42	0%	10%	24%	31%	36%
Hispanic	449	3%	6%	16%	38%	37%
African American	116	2%	6%	16%	52%	25%
White	309	3%	6%	19%	37%	34%
Ethnicity unknown	209	3%	8%	20%	37%	33%

* Results for groups with fewer than 11 members are not reported.

Table 7.B.3 Demographic Summary for Science, All Examinees

		Percentage in Performance Level				
	Number Tested	Far Below Basic	Below Basic	Basic	Proficient	Advanced
All valid scores	13,303	3%	7%	26%	41%	24%
Male	8,496	3%	7%	25%	41%	25%
Female	4,780	3%	7%	26%	42%	22%
Gender Unknown	27	0%	11%	33%	41%	15%
American Indian	108	2%	6%	21%	43%	29%
Asian American	845	3%	9%	31%	39%	18%
Pacific Islander	65	2%	8%	28%	42%	22%
Filipino	366	3%	8%	31%	42%	15%
Hispanic	6,879	3%	7%	24%	42%	25%
African American	1,346	3%	7%	23%	42%	25%
White	3,310	2%	6%	27%	41%	24%
Ethnicity unknown	384	3%	7%	30%	39%	22%
English Only	8,160	3%	7%	26%	41%	23%
Initially–Fluent English Proficient	237	4%	12%	31%	36%	17%
English Learner	4,218	3%	7%	24%	43%	24%
Reclassified–Fluent English Proficient	538	2%	6%	24%	42%	26%
English Proficient unknown	150	1%	6%	24%	45%	24%
Mental Retardation	5,822	2%	6%	27%	43%	23%
Hard of Hearing	99	1%	8%	23%	41%	26%
Deafness	104	0%	7%	17%	54%	22%
Speech/Language Impairment	298	0%	0%	11%	56%	33%
Visual Impairment	148	9%	13%	25%	30%	22%
Emotional Disturbance	102	0%	2%	11%	47%	40%
Orthopedic Impairment	1,213	8%	15%	26%	31%	20%
Other Health Impairment	537	1%	4%	19%	47%	29%
Specific Learning Impairment	839	0%	0%	6%	47%	47%
Deaf Blindness	16	25%	13%	38%	19%	6%
Multiple Group	621	10%	15%	29%	31%	16%
Autism	3,280	2%	7%	30%	41%	20%
Traumatic Brain Injury	86	8%	3%	16%	50%	22%
Unknown	138	1%	7%	26%	45%	21%
Not Econ. Disadvantaged	4,405	3%	8%	30%	38%	20%
Economically Disadvantaged	8,630	2%	6%	23%	43%	25%
Unknown Economic Status	268	3%	4%	22%	47%	24%
Primary Ethnicity—Not Economically Disadvantaged						
American Indian	41	2%	5%	29%	29%	34%
Asian American	428	4%	9%	33%	37%	17%
Pacific Islander	22	0%	18%	27%	32%	23%
Filipino	220	3%	10%	32%	40%	15%
Hispanic	1,207	5%	10%	27%	38%	21%
African American	374	5%	9%	30%	36%	20%
White	1,955	2%	7%	30%	40%	21%
Ethnicity unknown	158	4%	8%	31%	35%	23%

Primary Ethnicity—Economically Disadvantaged						
American Indian	66	2%	6%	17%	50%	26%
Asian American	398	3%	9%	28%	41%	19%
Pacific Islander	42	2%	2%	29%	45%	21%
Filipino	138	3%	7%	30%	46%	15%
Hispanic	5,570	2%	6%	24%	42%	25%
African American	946	3%	6%	21%	44%	26%
White	1,292	1%	6%	22%	43%	29%
Ethnicity unknown	178	2%	7%	29%	40%	22%
Primary Ethnicity—Unknown Economic Status						
American Indian	0	–	–	–	–	–
Asian American	19	11%	11%	26%	37%	16%
Pacific Islander	0	–	–	–	–	–
Filipino	0	–	–	–	–	–
Hispanic	102	1%	3%	21%	54%	22%
African American	26	0%	8%	15%	38%	38%
White	63	5%	2%	19%	40%	35%
Ethnicity unknown	48	4%	4%	29%	48%	15%

* Results for groups with fewer than 11 members are not reported.

Appendix 7.C—Type of Score Report

Table 7.C.1 Score Reports Reflecting CAPA Results

2011 STAR CAPA Student Reports	
Description	**Distribution**
The CAPA Student Report	
This report provides parents/guardians and teachers with the student's results, presented in tables and graphs. Data presented include the following: • Scale scores • Performance levels (advanced, proficient, basic, below basic, and far below basic)	This report includes individual student results and is not distributed beyond parents/guardians and the student's school. Two copies of this report are provided for each student. One is for the student's current teacher and one is distributed by the school district to parents/guardians.
Student Record Label	
These reports are printed on adhesive labels to be affixed to the student's permanent school records. Each student shall have an individual record of accomplishment that includes STAR testing results (see California *EC* Section 60607[a]). Data presented include the following for each content area tested: • Scale scores • Performance levels	This report includes individual student results and is not distributed beyond the student's school.
Student Master List	
This report is an alphabetical roster that presents individual student results. It primarily includes the following data: • Scale scores • Performance levels	This report provides administrators and teachers with all students' results within each grade or within each grade and year-round schedule at a school. Because this report includes individual student results, it is not distributed beyond the student's school.
Student Master List Summary	
This report summarizes student results at the school, district, county, and state levels for each grade. It does *not* include any individual student information. For each CAPA grade and level, the following data are summarized by content area tested: • Number of students enrolled • Number and percent of students tested • Number and percent of valid scores • Number tested with scores • Mean scale score • Scale score standard deviation • Number and percent of students scoring at each performance level	This report is a resource for evaluators, researchers, teachers, parents/guardians, community members, and administrators. One copy is packaged for the school and one for the school district. This report is also produced for school districts, counties, and the state. ***Note:*** The data in this report may be shared with parents/guardians, community members, and the media only if the data are for 11 or more students.

2011 STAR CAPA Student Reports	
Description	**Distribution**
Subgroup Summary	
This set of reports disaggregates and reports results by the following subgroups: • All students • Disability status (Disabilities among CAPA students include specific disabilities.) • Economic status • Gender • English proficiency • Primary ethnicity These reports contain no individual student-identifying information and are aggregated at the school, district, county, and state levels. CAPA statistics are listed by CAPA level. For each subgroup within a report and for the total number of students, the following data are included: • Total number tested in the subgroup • Percent tested in the subgroup as a percent of all students tested • Number and percent of valid scores • Number tested who received scores • Mean scale score • Standard deviation of scale score • Number and percent of students scoring at each performance level	This report is a resource for evaluators, researchers, teachers, parents/guardians, community members, and administrators. One copy is packaged for the school and one for the school district. This report is also produced for school districts, counties, and the state. ***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.

2011 STAR CAPA Student Reports	
Description	**Distribution**
Subgroup Summary—Ethnicity for Economic Status	
This report, a part of the Subgroup Summary, disaggregates and reports results by cross-referencing each ethnicity with economic status. The economic status for each student is "economically disadvantaged," "not economically disadvantaged," or "economic status unknown." A student is defined as "economically disadvantaged" if the most educated parent of the student, as indicated in the answer document or Pre-ID, has not received a high school diploma or the student is eligible to participate in the free or reduced-price lunch program also known as the National School Lunch Program (NSLP). As with the standard Subgroup Summary, this disaggregation contains no individual student-identifying information and is aggregated at the school, district, county, and state levels. CAPA statistics are listed by CAPA level. For each subgroup within a report, and for the total number of students, the following data are included: • Total number tested in the subgroup • Percent tested in the subgroup as a percent of all students tested • Number and percent of valid scores • Number tested who received scores • Mean scale score • Standard deviation of scale score • Number and percent of students scoring at each performance level	This report is a resource for evaluators, researchers, teachers, parents/guardians, community members, and administrators. One copy is packaged for the school and one for the school district. This report is also produced for school districts, counties, and the state. ***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.

Chapter 8: Analyses

This chapter summarizes the task (item)- and test-level statistics obtained for the CAPA administered during the spring of 2011.

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

1. Classical Item Analyses
2. Reliability Analyses
3. Analyses in Support of Validity Evidence
4. Item Response Theory (IRT) Analyses
5. Differential Item Functioning (DIF) Analyses

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

1. Appendix 8.A on page 88 presents the classical item analyses including average item score (AIS) and polyserial correlation coefficient, and associated flags, for the operational and field-test tasks of each test. Also presented in this appendix is information about the distribution of scores for the operational tasks. In addition, the mean, minimum, and maximum of AIS and polyserial correlation for each operational task are presented in Table 8.2 on page 70.
2. Appendix 8.B on page 105 presents results of the reliability analyses of total test scores for the population as a whole and for selected subgroups. Also presented are results of the analyses of the accuracy and consistency of the performance classifications.
3. Appendix 8.C on page 119 presents tables showing the correlations between scores obtained on the CAPA measured in the different content areas, which are provided as an example of the evidence of the validity of the interpretation and uses of CAPA scores. The results for the overall test population are presented in Table 8.4; the tables in Appendix 8.C summarize the results for various subgroups. Also included in Appendix 8.C are results of the rater agreement for each operational task.
4. Appendix 8.D on page 138 presents the results of IRT analyses including the distribution of tasks based on their fit to the Rasch model and the summaries of Rasch item difficulty statistics (*b*-values) for the operational and field-test tasks. In addition, the appendix presents the scoring tables obtained as a result of the IRT equating process. Information related to the evaluation of linking tasks is presented in Table 8.5 on page 82; these linking tasks were used in the equating process discussed later in this chapter.
5. Appendix 8.E on page 154 presents the results of the DIF analyses applied to all operational and field-test tasks for which sufficient student samples were available. In this appendix, tasks flagged for significant DIF are listed. Also given are the distributions of items across DIF categories.

Samples Used for the Analyses

CAPA analyses were conducted at different times after test administration and involved varying proportions of the full CAPA data. The IRT results presented in Appendix 8.D are based on the equating sample that includes all valid cases available by early June 2011. The classical item analyses presented in Table 8.2, Appendix 8.A, and item-level DIF results presented in Appendix 8.E are based on the P1 data, which contained more than 99 percent

of the test results of the entire test-taking population. All other analyses for this technical report, including reliability statistics in Appendix 8.B and the content area correlations presented in Appendix 8.C, were calculated using the P2 data file. Summary statistics describing the samples are presented in Table 8.1; the samples used to generate scoring tables are labeled as "Equating Samples."

Table 8.1 CAPA Raw Score Means and Standard Deviations: Total P2 Population and Equating Sample

Content Area	Level	P2			Equating Sample			
		N	Mean	SD	N	% of P2	Mean	SD
English–Language Arts	I	13,719	25.69	11.16	5,356	39	26.00	10.80
	II	6,643	20.00	6.14	2,960	45	20.34	6.08
	III	7,112	22.78	6.24	3,124	44	23.05	6.15
	IV	9,858	20.24	6.65	4,379	44	20.61	6.57
	V	10,217	21.84	6.73	4,936	48	22.23	6.48
Mathematics	I	13,689	22.79	10.82	5,343	39	23.11	10.47
	II	6,624	20.34	6.28	2,947	44	20.53	6.22
	III	7,098	18.57	5.46	3,116	44	18.75	5.43
	IV	9,845	19.65	6.89	4,374	44	20.06	6.75
	V	10,196	20.04	7.04	4,931	48	20.31	6.93
Science	I	3,512	23.23	11.72	1,329	38	23.58	11.28
	III	3,391	21.42	6.26	1,464	43	21.66	6.16
	IV	3,155	21.40	5.82	1,415	45	22.14	5.58
	V	3,245	20.59	5.99	1,581	49	20.97	5.77

Classical Analyses

Average Item Score

The Average Item Score (AIS) indicates the average score that students obtained on a task. Desired values generally fall within the range of 30 percent to 80 percent of the maximum obtainable task score. Occasionally, a task that falls outside this range is included in a test form because of the quality and educational importance of the task content or because it is the best available measure for students with very high or low achievement.

CAPA task scores range from 0 to 5 for Level I and 0 to 4 for Levels II through V. For tasks scored using a 0–4 point rubric, 30 percent is represented by the value 1.20 and 80 percent is represented by the value 3.20. For tasks scored using a 0–5 point rubric, 30 percent is represented by the value 1.50 and 80 percent is represented by the value 4.00.

Polyserial Correlation of the Task Score with the Total Test Score

This statistic describes the relationship between students' scores on a specific task and their total test scores. The polyserial correlation is used when an interval variable is correlated with an ordinal variable that is assumed to reflect an underlying continuous latent variable.

Polyserial correlations are based on a polyserial regression model (Drasgow, 1988). The ETS proprietary software Generalized Analysis System (GENASYS) estimates the value of β for each item using maximum likelihood. In turn, it uses this estimate of β to compute the polyserial correlation from the following formula:

$$r_{polyreg} = \frac{\hat{\beta} s_{tot}}{\sqrt{\hat{\beta}^2 s^2_{tot} + 1}} \tag{8.1}$$

where,

s_{tot} is the standard deviation of the students' total scores; and

β is the item parameter to be estimated from the data, with the estimate denoted as $\hat{\beta}$, using maximum likelihood.

β is a regression coefficient (slope) for predicting the continuous version of a binary item score onto the continuous version of the total score. There are as many regressions as there are boundaries between scores with all sharing a common slope, β. For a polytomously scored item, there are k-1 regressions, where k is the number of score points on the item. Beta (β) is the slope for all k-1 regressions.

The polyserial correlation is sometimes referred to as a discrimination index because it is an indicator of the degree to which students who do well on the total test also do well on a given task. A task is considered discriminating if high-ability students tend to receive higher scores and low-ability students tend to receive lower scores on the task.

Tasks with negative or extremely low correlations can indicate serious problems with the task itself or can indicate that students have not been taught the content. Based on the range of polyserials produced in field test analyses, an indicator of poor discrimination was set to less than 0.60.

A descriptive summary of the classical item statistics for the overall test are presented in Table 8.2. The task-by-task values are presented in Table 8.A.1 through Table 8.A.14. Some tasks were flagged for unusual statistics; these flags are shown in the tables. Although the flag definition appears in the heading of each table, the flags are displayed in the body of the tables only where applicable for the specific CAPA presented. The flag classifications are as follows:

- Difficulty flags
  - A: Low average task score (below 1.5 at Level I; below 1.2 at Levels II–V)
  - H: High average task score (above 4.0 at Level I; above 3.2 at Levels II–V)
- Discrimination flag
  - R: Polyserial correlation less than .60
- Omit/nonresponse/flag
  - O: Omit/nonresponse rates greater than 5 percent

Table 8.2 Average Item Score and Polyserial Correlation

Content Area	Level	No. of items	No. of Examinees	Mean		Minimum		Maximum	
				AIS	Polyserial	AIS	Polyserial	AIS	Polyserial
ELA	I	8	13,716	3.21	0.79	2.73	0.71	4.24	0.84
	II	8	6,627	2.51	0.72	1.91	0.60	3.57	0.80
	III	8	7,099	2.86	0.78	2.21	0.68	3.63	0.85
	IV	8	9,858	2.54	0.77	1.64	0.71	3.49	0.84
	V	8	10,214	2.74	0.80	1.97	0.67	3.50	0.89
Mathematics	I	8	13,686	2.85	0.75	2.66	0.72	3.23	0.78
	II	8	6,608	2.55	0.73	1.38	0.59	3.49	0.84
	III	8	7,085	2.32	0.68	1.76	0.62	2.98	0.80
	IV	8	9,845	2.47	0.75	1.83	0.61	3.00	0.85
	V	8	10,193	2.51	0.76	2.01	0.72	3.35	0.81

Content Area	Level	No. of items	No. of Examinees	Mean		Minimum		Maximum	
				AIS	Polyserial	AIS	Polyserial	AIS	Polyserial
Science	I	8	3,512	2.91	0.80	2.65	0.76	3.50	0.83
	III	8	3,385	2.69	0.73	2.34	0.61	2.94	0.80
	IV	8	3,155	2.69	0.73	2.34	0.67	2.96	0.75
	V	8	3,245	2.58	0.76	2.14	0.69	3.33	0.81

As noted previously, the score distributions for individual operational tasks comprising each CAPA test are provided by content area and level in Table 8.A.15 through Table 8.A.17.

Reliability Analyses

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

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

Reliability coefficients may range from 0 to 1. The higher the reliability coefficient for a set of scores, the more likely individuals would be to obtain very similar scores if they were retested. The formula for the internal consistency reliability as measured by Cronbach's Alpha (Cronbach, 1951) is defined by equation 8.2:

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n} s_i^2}{s_t^2}\right] \tag{8.2}$$

where,

n is the number of tasks,

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

s_t^2 is the variance of the total score .

The standard error of measurement (SEM) provides a measure of score instability in the score metric. The SEM is defined by:

$$s_e = s_t\sqrt{1-\alpha} \tag{8.3}$$

where,

α is the reliability estimated using equation 8.2, and

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

The SEM is computed by substituting sample estimates of the population variances in the defining formulas (equations 8.3 and 8.4). The SEM is particularly useful in determining the confidence interval (CI) that captures an examinee's true score. Assuming that measurement error is normally distributed, it can be said that upon infinite replications of the testing occasion, approximately 95 percent of the CIs of ±1.96 SEM around the observed score would contain an examinee's true score (Crocker & Algina, 1986). For example, if an examinee's observed score on a given test equals 15 points, and the SEM equals 1.92, one can be 95 percent confident that the examinee's true score lies between 11 and 19 points (15 ± 3.76 rounded to the nearest integer).

Table 8.3 gives the reliability and SEM for CAPA tests along with the number of items and examinees upon which those analyses were performed.

Table 8.3 Reliabilities and Standard Errors of Measurement for the CAPA

Content Area	Level	No. of Items	No. of Examinees	Reliab.	Scale Score Mean	Scale Score S.D.	Scale Score SEM	Raw Score Mean	Raw Score S.D.	Raw Score SEM
English–Language Arts	I	8	13,719	0.88	40.79	10.63	3.68	25.69	11.16	3.86
	II	8	6,643	0.82	38.72	5.70	2.40	20.00	6.14	2.59
	III	8	7,112	0.88	39.53	6.17	2.17	22.78	6.24	2.19
	IV	8	9,858	0.86	39.20	7.44	2.78	20.24	6.65	2.48
	V	8	10,217	0.89	38.88	6.44	2.09	21.84	6.73	2.19
Mathematics	I	8	13,689	0.84	36.08	8.70	3.43	22.79	10.82	4.27
	II	8	6,624	0.82	37.45	7.64	3.22	20.34	6.28	2.64
	III	8	7,098	0.77	36.33	4.99	2.41	18.57	5.46	2.63
	IV	8	9,845	0.85	37.26	8.11	3.11	19.65	6.89	2.64
	V	8	10,196	0.85	37.32	7.63	2.96	20.04	7.04	2.73
Science	I	8	3,512	0.89	36.20	10.67	3.55	23.23	11.72	3.90
	III	8	3,391	0.84	36.42	5.27	2.09	21.42	6.26	2.49
	IV	8	3,155	0.84	36.23	5.72	2.28	21.40	5.82	2.32
	V	8	3,245	0.86	35.82	5.08	1.92	20.59	5.99	2.26

Subgroup Reliabilities and SEMs

The reliabilities of the CAPA were examined for various subgroups of the examinee population. The subgroups included in these analyses were defined by their gender, ethnicity, economic status, disability group, and English-language fluency. As of 2008, reliability analyses are also presented by primary ethnicity within economic status.

Table 8.B.1 through Table 8.B.6 present the reliabilities and SEM information for the total test scores for each subgroup. Note that the reliabilities are reported only for samples that are comprised of 11 or more examinees. Also, in some cases, score reliabilities were not estimable and are presented in the tables as hyphens.

Conditional Standard Errors of Measurement

As part of the IRT-based equating procedures, scale-score conversion tables and conditional standard errors of measurement (CSEMs) are produced. CSEMs for CAPA scale scores are based on IRT and are calculated by the IRTEQUATE module in GENASYS.

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

$$\mathrm{CSEM}(\hat{\theta}) = \frac{1}{\sqrt{\mathrm{I}(\theta)}} a \tag{8.4}$$

where,

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

I(θ) is the test information function at ability level θ.

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

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

Table 8.D.10 through Table 8.D.23 in Appendix 8.D present the scale score CSEMs at the score required for a student to be classified in the below basic, basic, proficient, and advanced performance levels for the CAPA. The pattern of lower values of CSEMs at the basic and proficient levels are expected since (1) more items tend to be of middle difficulty; and (2) items at the extremes still provide information toward the middle of the scale. This results in more precise scores in the middle of the scale and less precise scores in the extremes of the scale.

Decision Classification Analyses

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

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

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

In each case, the proportion of classifications with exact agreement is the sum of the entries in the diagonal of the contingency table representing the multivariate distribution. Reliability of classification at a cut score is estimated by collapsing the multivariate distribution at the

passing score boundary into an *n* by *n* table (where *n* is the number of performance levels) and summing the entries in the diagonal. Figure 8.1 and Figure 8.2 present the two scenarios graphically.

Figure 8.1 Decision Accuracy for Achieving a Performance Level

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

Figure 8.2 Decision Consistency for Achieving a Performance Level

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

The results of these analyses are presented in Table 8.B.7 through Table 8.B.20 in Appendix 8.B starting on page 111.

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

Results for classifications collapsed to below-proficient versus proficient and above, which are the critical categories for Adequate Yearly Progress (AYP) calculations, are also presented in the tables.

Validity Evidence

Validity refers to the degree to which each interpretation or use of a test score is supported by evidence that is gathered (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999; ETS, 2002). It is a central concern underlying the development, administration, and scoring of a test and the uses and interpretations of test scores.

Validation is the process of accumulating evidence to support each proposed score interpretation or use. It involves more than a single study or gathering one particular kind of evidence. Validation involves multiple investigations and various kinds of evidence (AERA, APA, & NCME, 1999; Cronbach, 1971; ETS, 2002; Kane, 2006). The process begins with test design and continues through the entire assessment process, including item development and field testing, analyses of item and test data, test scaling, scoring, and score reporting.

This section presents the evidence gathered to support the intended uses and interpretations of scores for the CAPA testing program. The description is organized in the

manner prescribed by *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999). These standards require a clear definition of the purpose of the test, which includes a description of the qualities—called constructs—that are to be assessed by a test, the population to be assessed, as well as how the scores are to be interpreted and used.

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

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

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

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

Purpose of the CAPA

As mentioned in Chapter 1, the CAPA is used in calculating school and district API. Additionally, the CAPA results for ELA and Mathematics in grades two through eight and grade ten are used in determining AYP that applies toward meeting the requirement of the Elementary and Secondary Education Act (ESEA), which is to have all students score at proficient or above by 2014.

The Constructs to Be Measured

The CAPA is designed to show how well students with an IEP and who have significant cognitive disabilities perform relative to the California content standards. These content standards were approved by the SBE; they describe what students should know and be able to do at each level.

Test blueprints and specifications written to define the procedures used to measure the content standards provide an operational definition of the construct to which each set of standards refers—that is, they define, for each content area to be assessed, the tasks to be presented, the administration instructions to be given, and the rules used to score examinee responses. They control as many aspects of the measurement procedure as possible so that the testing conditions will remain the same over test administrations (Cronbach, 1971; Cronbach, Gleser, Nanda, & Rajaratnam, 1972) to minimize construct-irrelevant score variance (Messick, 1989). The test blueprints for the CAPA can be found on the CDE STAR CAPA Blueprints Web page at http://www.cde.ca.gov/ta/tg/sr/capablueprints.asp. ETS has developed all CAPA test tasks to conform to the SBE-approved content standards and test blueprints.

Interpretations and Uses of the Scores Generated

Total scores expressed as scale scores, and student performance levels are generated for each grade-level test. On the basis of a student's total score, an inference is drawn about how much knowledge and skill in the content area the student has acquired. The total score

also is used to classify students in terms of their level of knowledge and skill in the content area. The classifications are called performance levels and are labeled as follows: advanced, proficient, basic, below basic, and far below basic.

The tests that make up the STAR Program, along with other assessments, provide results or score summaries that are used for different purposes. The four major purposes are:

1. Communicating with parents and guardians;
2. Informing decisions needed to support student achievement;
3. Evaluating school programs; and
4. Providing data for state and federal accountability programs for schools.

These are the only uses and interpretations of scores for which validity evidence has been gathered. If the user wishes to interpret or use the scores in other ways, the user is cautioned that the validity of doing so has not been established (AERA, APA, & NCME, 1999, Standard 1.3). The user is advised to gather evidence to support these additional interpretations or uses (AERA, APA, & NCME, 1999, Standard, 1.4).

Intended Test Population(s)

Students with an IEP and who have significant cognitive disabilities in grades two through eleven take the CAPA when they are unable to take the CSTs with or without accommodations or modifications or the CMA with accommodations. Participation in the CAPA and eligibility are determined by a student's IEP team. Only students whose parents/guardians have submitted written requests to exempt them from STAR Program testing do not take the tests.

Validity Evidence Collected

Evidence Based on Content

According to *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), analyses that demonstrate a strong relationship between a test's content and the construct that the test was designed to measure can provide important evidence of validity. In current K–12 testing, the construct of interest usually is operationally defined by state content standards and the test blueprints that specify the content, format, and scoring of items that are admissible measures of the knowledge and skills described in the content standards. Evidence that the items meet these specifications and represent the domain of knowledge and skills referenced by the standards supports the inference that students' scores on these items can appropriately be regarded as measures of the intended construct.

As noted in the AERA, APA, and NCME *Test Standards* (1999), evidence based on test content may involve logical analyses of test content in which experts judge the adequacy with which the test content conforms to the test specifications and represents the intended domain of content. Such reviews can also be used to determine whether the test content contains material that is not relevant to the construct of interest. Analyses of test content may also involve the use of empirical evidence of item quality.

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

Description of the state standards—As was noted in Chapter 1, the SBE adopted rigorous content standards in 1997 and 1998 in four major content areas: ELA, history–social science, mathematics, and science. These standards were designed to guide instruction and learning for all students in the state and to bring California students to world-class levels of achievement.

Specifications and blueprints—ETS maintains task specifications for the CAPA. The task specifications describe the characteristics of the tasks that should be written to measure each content standard. A thorough description of the specifications can be found in Chapter 3, starting on page 16. Once the tasks are developed and field-tested, ETS selects all CAPA test tasks to conform to the SBE-approved California content standards and test blueprints. Test blueprints for the CAPA were proposed by ETS and reviewed and approved by the Assessment Review Panels (ARPs), which are advisory panels to the CDE and ETS on areas related to task development for the CAPA. Tasks were also reviewed and approved by the CDE and presented to the SBE for adoption. There have been no recent changes in the blueprints for the CAPA; the blueprints were most recently revised and adopted by the SBE in 2006 for implementation beginning in 2008. The test blueprints for the CAPA can be found on the CDE STAR CAPA Blueprints Web page at http://www.cde.ca.gov/ta/tg/sr/capablueprints.asp.

Task development process—A detailed description of the task development process for the CAPA is presented in Chapter 3, starting on page 16.

Task review process—Chapter 3 explains in detail the extensive item review process applied to tasks written for use in the CAPA. In brief, tasks written for the CAPA undergo multiple review cycles and involve multiple groups of reviewers. One of the reviews is carried out by an external reviewer, that is, the ARPs. The ARPs are responsible for reviewing all newly developed tasks for alignment to the California content standards.

Form construction process—For each test, the content standards, blueprints, and test specifications are used as the basis for choosing tasks. Additional targets for item difficulty and discrimination that are used for test construction were defined in light of what are desirable statistical characteristics in test tasks and statistical evaluations of the CAPA tasks.

Guidelines for test construction were established with the goal of maintaining parallel forms to the greatest extent possible from year to year. Details can be found in Chapter 4, starting on page 26.

Additionally, an external review panel, the Statewide Pupil Assessment Review (SPAR), is responsible for reviewing and approving the achievement tests to be used statewide for the testing of students in California public schools, grades two through eleven. More information about the SPAR is given in Chapter 3, starting on page 22.

Alignment study—Strong alignment between standards and assessments is fundamental to meaningful measurement of student achievement and instructional effectiveness. Alignment results should demonstrate that the assessments represent the full range of the content standards and that these assessments measure student knowledge in the same manner and at the same level of complexity as expected in the content standards.

Human Resource Research Organization (HumRRo) performed an alignment study for the CAPA in April 2007. The result of this study was a report entitled *Independent*

Evaluation of the Alignment of the California Standards Tests (CSTs) and the California Alternate Performance Assessment (CAPA).

HumRRO utilized the Webb alignment method to evaluate the alignment of the performance tasks field-tested in the 2007 CAPA to the California content standards. The Webb method requires a set of raters to evaluate each test item on two different dimensions: (1) the standard(s) targeted by items, and (2) the depth of knowledge required of students to respond to items. These ratings form the basis of the four separate Webb alignment analyses: categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance-of-knowledge representation. The results indicated that the performance tasks assess the majority of CAPA standards well across levels for both ELA and mathematics. A copy of the study is available as a CDE Web document at http://www.cde.ca.gov/ta/tg/sr/documents/alignmentreport.pdf.

Evidence Based on Relations to Other Variables

Empirical results concerning the relationships between the scores on a test and measures of other variables external to the test can also provide evidence of validity when these relationships are found to be consistent with the definition of the construct that the test is intended to measure. As indicated in the *Test Standards* (AERA, APA, & NCME, 1999), the variables investigated can include other tests that measure the same construct and different constructs, criterion measures that scores on the test are expected to predict, as well as demographic characteristics of examinees that are expected to be related and unrelated to test performance.

Differential Item Functioning Analyses

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

The results of the DIF analyses are presented in Appendix 8.E, which starts on page 154. The vast majority of the tasks exhibited little or no significant DIF, suggesting that, in general, scores based on the CAPA tasks would have the same meaning for individuals who differed in their demographic characteristics.

Correlations Between Content Areas

To the degree that students' content-area scores correlate as expected, evidence of the validity in regarding those scores as measures of the intended constructs is provided. Table 8.4 gives the correlations between scores on the CAPA content-area tests, means and standard deviations of total raw scores, and the numbers of students on which these correlations were based. Sample sizes for individual tests are shown on the diagonals of the correlation matrices, and the numbers of students on which the correlations were based are shown on the lower off-diagonals. The correlations are provided in the upper off-diagonals.

At Level I, the correlations between students' ELA, mathematics, and science scores were high. For Levels II and above, the correlations between content area scores tended to be more moderate.

Table 8.C.1 through Table 8.C.35 in Appendix 8.C provide the content-area correlations by gender, ethnicity, English-language fluency, economic status, and disability. Similar patterns of correlations between students' ELA, mathematics, and science scores were found within the subgroups.

Note that the correlations are reported only for samples that are comprised of more than 10 examinees. Correlations between any two content areas where 10 or fewer examinees took the tests are expressed as hyphens. Correlations between two content area tests that cannot be administered to the same group of students are expressed as "N/A."

Table 8.4 CAPA Content-area Correlations for CAPA Levels

Level	Content	ELA	Mathematics	Science
I	ELA	**13,719**	0.79	0.82
	Mathematics	13,680	**13,689**	0.78
	Science	3,509	3,510	**3,512**
II	ELA	**6,643**	0.74	N/A
	Mathematics	6,622	**6,624**	N/A
	Science	N/A	N/A	N/A
III	ELA	**7,112**	0.72	0.69
	Mathematics	7,095	**7,098**	0.68
	Science	3,389	3,388	**3,391**
IV	ELA	**9,858**	0.76	0.72
	Mathematics	9,838	**9,845**	0.72
	Science	3,154	3,151	**3,155**
V	ELA	**10,217**	0.70	0.69
	Mathematics	10,180	**10,196**	0.70
	Science	3,239	3,240	**3,245**

Evidence Based on Response Processes

As noted in the APA, AERA, and NCME *Standards* (1999), additional support for a particular score interpretation or use can be provided by theoretical and empirical evidence indicating that examinees are using the intended response processes when responding to the items in a test. This evidence may be gathered from interacting with examinees in order to understand what processes underlie their item responses. Finally, evidence may also be derived from feedback provided by observers or judges involved in the scoring of examinee responses.

Evidence of Interrater Agreement

Rater consistency is critical to the scores of CAPA tasks and their interpretations. These findings provide evidence of the degree to which raters agree in their observations about the qualities evident in students' responses. In order to monitor and evaluate the accuracy of rating, approximately 10 percent of students' test responses were scored twice. They were scored once by the primary examiner (rater 1) and a second time by an independent, trained observer (rater 2). Evidence that the raters' scores are consistent helps to support the inference that the scores have the intended meaning. The data collected were used to evaluate interrater agreement.

Interrater Agreement

As noted previously, approximately 10 percent of the test population's responses to the tasks were scored by two raters. The range of percentage of students for whom the raters were in exact agreement was 89 to 97 percent in ELA, 93 to 99 percent in mathematics, and 90 to 98 percent in science for Level V.

The findings for each operational task for each test are presented in Table 8.C.36 through Table 8.C.40 in Appendix 8.C, which start on page 133.

Evidence Based on Internal Structure

As suggested by the *Standards* (AERA, APA, & NCME, 1999), evidence of validity can also be obtained from studies of the properties of the item (task) scores and the relationship between these scores and scores on components of the test. To the extent that the score properties and relationships found are consistent with the definition of the construct measured by test, support is gained for interpreting these scores as measures of the construct.

For the CAPA, it is assumed that a single construct underlies the total scores obtained on each test. Evidence to support this assumption can be gathered from the results of task analyses, evaluations of internal consistency, and studies of model-data fit and reliability.

Reliability

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

Overall reliability—The reliability analyses are presented in Table 8.3. The results indicate that the reliabilities for all CAPA levels for ELA, mathematics, and science tended to be high, ranging from 0.77 to 0.89.

Subgroup reliabilities—The reliabilities of the operational CAPA scores were also examined for various subgroups of the examinee population that differed in their demographic characteristics. The characteristics considered were gender, ethnicity, economic status, disability group, English-language fluency, and ethnicity-by-economic status. The results of these analyses can be found in Table 8.B.1 though Table 8.B.6.

Evidence Based on Consequences of Testing

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

IRT Analyses

The IRT model used to calibrate the CAPA test tasks is the one-parameter partial credit (1PPC) model, a more restrictive version of the generalized partial-credit model (Muraki, 1992), in which all tasks are assumed to be equally discriminating. This model states that the probability that an examinee with ability θ will perform in the *kth* category of m_j ordered score categories of task *j* can be expressed as:

$$P_{jk}(\theta) = \frac{\exp\left[\sum_{v=1}^{k} 1.7a_j(\theta - b_j - d_{jv})\right]}{\sum_{c=1}^{m_j} \exp\left[\sum_{v=1}^{c} 1.7a_j(\theta - b_j - d_{jv})\right]} \tag{8.5}$$

where,

m_j is the number of possible score categories (c=1…m_j) for task *j,*

a_j is the slope parameter (equal to 0.588) for task *j*,

b_j is the difficulty of task *j*, and

d_{jv} is the threshold parameter for category v of task *j*.

For the task calibrations, the PARSCALE program (Muraki & Bock, 1995) was constrained by setting a common discrimination value for all tasks equal to 1.0 / 1.7 (or 0.588) and by setting the lower asymptote for all tasks to zero. The resulting estimation is equivalent to the Rasch partial credit model for polytomously scored tasks.

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

The parameters estimated for each task were evaluated for model-data fit, as described below.

IRT Model-Data Fit Analyses

ETS psychometricians classify operational and field-test tasks for the CAPA into discrete categories based on an evaluation of how well each task was fit by the Rasch partial credit 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.

Flag A

- 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

- 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

- 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

- 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

- 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, and the ratings of F indicate a poor model fit.

Model-fit Assessment Results

The model-fit assessment is performed twice in the administration cycle. The assessment is first performed before scoring tables are produced and released. The assessment is performed again as part of the final item analyses when much larger samples are available. The flags produced as a result of this assessment are placed in the item bank. The test developers are asked to avoid the items flagged as D, 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 and by CDE content specialists.

The number of the operational and field-test tasks in each IRT model data fit classification is presented in Table 8.D.1 through Table 8.D.6, which start on page 138.

Evaluation of Scaling

Calibrations of the 2011 forms were scaled to the previously obtained reference scale estimates in the item bank using the Stocking and Lord (1983) procedure. Details on the scaling procedures are provided on page 13 of Chapter 2.

The linking process is carried out iteratively by inspecting differences between the transformed new and old (reference) estimates for the linking items and removing items for which the item difficulty estimates changed significantly. Items with large weighted root-mean-square differences (WRMSDs) between item characteristic curves (ICCs) on the basis of the old and new difficulty estimates are removed from the linking set. Based on established procedures, any linking items for which the WRMSD was greater than 0.625 for Level I and 0.500 for Levels II through V were eliminated. This criterion has produced reasonable results over time in similar equating work done with other testing programs at ETS. For the 2011 CAPA tests, no linking tasks were eliminated.

Table 8.5 presents, for each CAPA, the number of linking tasks between the 2011 (new) form and the test form to which it was linked (2010); the number of tasks removed from the linking task sets; the correlation between the final set of new and reference difficulty estimates for the linking tasks; and the average WRMSD statistic across the final set of linking tasks.

Table 8.5 Evaluation of Common Items Between New and Reference Test Forms

Content Area	Level	No. Linking Tasks	Linking Tasks Removed	Final Correlation	WRMSD*
English–Language Arts	I	5	0	0.99	0.04
	II	5	0	1.00	0.03
	III	5	0	0.99	0.05
	IV	5	0	1.00	0.06
	V	5	0	0.99	0.05

Content Area	Level	No. Linking Tasks	Linking Tasks Removed	Final Correlation	WRMSD*
Mathematics	I	5	0	0.98	0.08
	II	5	0	1.00	0.06
	III	5	0	0.99	0.08
	IV	5	0	1.00	0.05
	V	5	0	0.96	0.11
Science	I	5	0	0.99	0.07
	III	5	0	0.99	0.06
	IV	5	0	0.96	0.07
	V	5	0	0.99	0.05

* Average over retained tasks

Summaries of Scaled IRT b-values

Once the IRT *b*-values are placed on the item bank scale, analyses are performed to assess the overall test difficulty and the distribution of tasks in a particular range of item difficulty.

Table 8.D.7 through Table 8.D.9 present univariate statistics (mean, standard deviation, minimum, and maximum) for the scaled IRT *b*-values. The results for the overall test are presented separately for the operational tasks and the field test tasks.

Post-scaling Results

As described on page 13 of Chapter 2, once the new item calibrations for each test are transformed to the base scale, using equation 2.2, transformed thetas were linearly converted to two-digit scale scores that ranged from 15 to 60. Complete raw-to-scale score conversion tables for the 2011 CAPA are presented in Table 8.D.10 through Table 8.D.23 in Appendix 8.D, starting on page 140. The raw scores and corresponding rounded converted scale scores are listed in those tables.

For all of the 2011 CAPA, scale scores were truncated at both ends of the scale so that the minimum reported scale score was 15 and the maximum reported scale score was 60. The scale scores defining the cut scores for all performance levels are presented in Table 2.2, which is on page 14 in Chapter 2.

Differential Item Functioning Analyses

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

DIF analyses were performed on all operational tasks and all tasks being field-tested for which sufficient student samples were available. The sample size requirements for the DIF analyses were 100 in the focal group and 400 in the combined focal and reference groups. These sample sizes were based on standard operating procedures with respect to DIF analyses at ETS.

DIF analyses of the polytomously scored CAPA tasks are completed using two procedures. The first is the Mantel-Haenszel (MH) ordinal procedure, which is based on the Mantel procedure (Mantel, 1963; Mantel & Haenszel, 1959). The MH ordinal procedure compares the proportion of examinees in the reference and focal groups obtaining each task score after matching the examinees on their total test score. As with dichotomously scored tasks, the common odds ratio is estimated across the matched score groups. The resulting estimate is interpreted as the relative likelihood of obtaining a given task score for members of two groups that are matched on ability.

As such, the common odds ratio provides an estimated effect size; a value of one indicates equal odds and thus no DIF (Dorans & Holland, 1993). The corresponding statistical test is H_0: $\alpha = 1$, where α is a common odds ratio assumed equal for all matched score categories $s = 1$ to S. Values of less than one indicate DIF in favor of the focal group; a value of one indicates the null condition; and a value greater than one indicates DIF in favor of the reference group. The associated ($MH\chi^2$) is distributed as a Chi-square random variable with one degree of freedom.

The $MH\chi^2$ Mantel Chi-square statistic is used in conjunction with a second procedure, the standardization procedure (Dorans & Schmitt, 1993). This procedure produces a DIF statistic based on the standardized mean difference (SMD) in average task scores between members of two groups that have been matched on their overall test score. The SMD compares the task means of the two studied groups after adjusting for differences in the distribution of members across the values of the matching variable (total test score).

The standardized mean difference is computed as the following:

$$SMD = \sum_m w_m \left(E_{fm} - E_{rm} \right) / \sum_m w_m \tag{8.6}$$

where,

$w_m / \sum w_m$ is the weighting factor at score level *m* supplied by the standardization group to weight differences in item performance between a focal group (E_{fm}) and a reference group (E_{rm}) (Doran & Kulick, 2006).

A negative SMD value means that, conditional on the matching variable, the focal group has a lower mean task score than the reference group. In contrast, a positive SMD value means that, conditional on the matching variable, the reference group has a lower mean task score than the focal group. The SMD is divided by the standard deviation (SD) of the total group task score in its original metric to produce an effect-size measure of differential performance.

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

The ETS classification system assigns tasks to one of the three DIF categories on the basis of a combination of statistical significance of the Mantel Chi-square statistic and the magnitude of the SMD effect-size:

DIF Category	Definition
A (negligible)	• The Mantel Chi-square statistic is not statistically significant (at the 0.05 level) or |SMD/SD| < 0.17.
B (moderate)	• The Mantel Chi-square statistic is statistically significant (at the 0.05 level) and 0.17 ≤ |SMD/SD| < 0.25.
C (large)	• The Mantel Chi-square statistic is statistically significant (at the 0.05 level) and |SMD/SD| > 0.25.

In addition, the categories identify which group is being advantaged; categories are displayed in Table 8.6. The categories have been used by all ETS testing programs for more than 13 years.

Table 8.6 DIF Flags Based on the ETS DIF Classification Scheme

Flag	Descriptor
A–	Negligible favoring members of the reference group
B–	Moderate favoring members of the reference group
C–	Large favoring members of the reference group
A+	Negligible favoring members of the focal group
B+	Moderate favoring members of the focal group
C+	Large favoring members of the focal group

Category C contains tasks with large values of DIF. As shown in Table 8.6, tasks classified as C+ tend to be easier for members of the focal group than for members of the reference group with comparable total scores. Tasks classified as C– tend to be more difficult for members of the focal group than for members of the reference group whose total scores on the test are like those of the focal group.

The results of the DIF analyses are presented in Appendix 8.E, which starts on page 154. Table 8.E.1 and Table 8.E.2 list the tasks exhibiting significant DIF. Test developers are instructed to avoid selecting field-test items flagged as having shown DIF that disadvantages a focal group (C-DIF) for future operational test forms unless their inclusion is deemed essential to meeting test-content specifications.

Table 8.7 lists specific subgroups that were used for DIF analyses for the CAPA.

Table 8.7 Subgroup Classification for DIF Analyses

DIF Type	Reference Group	Focal Group
Gender	Male	Female
Race/Ethnicity	White	• African American • American Indian • Asian • Combined Asian Group (Asian/Pacific Islander/Filipino) • Filipino • Hispanic/Latin American • Pacific Islander
Disability	Mental Retardation	• Autism • Deaf-Blindness • Deafness • Emotional Disturbance • Hard of Hearing • Multiple Disabilities • Orthopedic Impairment • Other Health Impairment • Specific Learning Disability • Speech or Language Impairment • Traumatic Brain Injury • Visual Impairment

Table 8.E.3 through Table 8.E.7 show the sample size for disability groups within test level and content area.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 1999. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory.* New York, NY: Holt.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16,* 292–334.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement (*2nd ed., pp. 443–507*).* Washington, D. C.: American Council on Education.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York, NY: Wiley.

Dorans, N. J. & Holland, P. W. (1993). *DIF detection and description: Mantel-Haenszel and standardization*. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp.35–66). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Dorans, N. J. & Kulick, E. (2006). Differential item functioning on the mini-mental state examination: An application of the Mantel-Haenszel and standardization procedures. *Medical Care*, *44,* 107–14.

Dorans, N. J. & Schmitt, A. P. (1993). Constructed response and differential item functioning: A pragmatic approach. In R.E. Bennett & W.C. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 135–65). Hillsdale, NH: Lawrence Erlbaum Associates, Inc.

Drasgow F. (1988). Polychoric and polyserial correlations. In L. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Vol. 7, pp. 69–74). New York: Wiley.

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

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Washington, DC: American Council on Education and National Council on Measurement in Education.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classification based on test scores. *Journal of Educational Measurement*, *32,* 179–97.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure, *Journal of the American Statistical Association*, *58,* 690–700.

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analyses of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22,* 719–48.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed. pp. 13–103). New York, NY: Macmillan.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16,* 159–76.

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

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

Appendix 8.A—Classical Analyses: Task Statistics

Table 8.A.1 AIS and Polyserial Correlation: Level I, ELA

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/ Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	1	4.24	.71	H
1	2	2.98	.63	
Operational	3	2.73	.79	
Operational	4	3.56	.82	
1	5	3.61	.59	R
Operational	6	2.82	.79	
Operational	7	3.46	.84	
1	8	3.37	.79	
Operational	9	2.95	.78	
Operational	10	3.02	.83	
1	11	3.81	.62	
Operational	12	2.87	.78	
2	2	3.37	.74	
2	5	4.09	.69	H
2	8	2.87	.71	
2	11	3.99	.78	
3	2	2.84	.69	
3	5	3.65	.56	R
3	8	3.09	.71	
3	11	3.62	.72	
4	2	3.16	.74	
4	5	3.03	.78	
4	8	3.06	.80	
4	11	3.99	.74	

Table 8.A.2 AIS and Polyserial Correlation: Level II, ELA

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/ Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	1	2.64	.74	
1	2	2.63	.72	
Operational	3	2.44	.72	
Operational	4	3.57	.71	H
1	5	1.92	.71	
Operational	6	2.67	.76	
Operational	7	2.21	.80	
1	8	2.06	.56	R
Operational	9	2.42	.67	
Operational	10	1.91	.60	R
1	11	3.38	.61	H
Operational	12	2.20	.77	
2	2	3.38	.58	R H
2	5	1.93	.71	
2	8	2.54	.65	
2	11	2.29	.65	
3	2	3.44	.61	H
3	5	3.48	.64	H
3	8	2.41	.59	R
3	11	2.60	.61	
4	2	2.47	.66	
4	5	3.48	.65	H
4	8	3.32	.54	R H
4	11	2.65	.61	

Table 8.A.3 AIS and Polyserial Correlation: Level III, ELA

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/ Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	1	3.63	.76	H
1	2	2.71	.68	
Operational	3	2.59	.84	
Operational	4	2.99	.76	
1	5	2.67	.81	
Operational	6	2.27	.68	
Operational	7	2.49	.84	
1	8	2.44	.77	
Operational	9	3.13	.76	
Operational	10	2.21	.85	
1	11	3.53	.70	H
Operational	12	3.53	.73	H
2	2	2.38	.78	
2	5	3.47	.68	H
2	8	3.42	.63	H
2	11	2.66	.59	R
3	2	2.45	.60	R
3	5	3.72	.64	H
3	8	3.47	.65	H
3	11	2.91	.56	R
4	2	2.58	.65	
4	5	2.65	.68	
4	8	2.92	.53	R
4	11	2.68	.60	

Table 8.A.4 AIS and Polyserial Correlation: Level IV, ELA

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/ Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	1	2.78	.75	
1	2	2.75	.71	
Operational	3	2.44	.79	
Operational	4	1.64	.74	
1	5	3.16	.74	
Operational	6	2.58	.81	
Operational	7	3.49	.71	H
1	8	2.87	.81	
Operational	9	2.62	.73	
Operational	10	2.59	.75	
1	11	2.90	.67	
Operational	12	2.16	.84	
2	2	2.69	.74	
2	5	3.03	.61	
2	8	2.44	.68	
2	11	1.66	.69	
3	2	2.86	.71	
3	5	2.79	.49	R
3	8	2.35	.53	R
3	11	1.73	.69	
4	2	2.67	.65	
4	5	2.52	.59	R
4	8	2.61	.70	
4	11	2.71	.68	

Table 8.A.5 AIS and Polyserial Correlation: Level V, ELA

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/ Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	1	3.12	.67	
1	2	2.90	.77	
Operational	3	2.56	.86	
Operational	4	2.77	.77	
1	5	2.90	.70	
Operational	6	1.97	.71	
Operational	7	3.15	.82	
1	8	2.52	.84	
Operational	9	2.43	.89	
Operational	10	2.43	.89	
1	11	3.27	.80	H
Operational	12	3.50	.79	H
2	2	2.77	.37	R
2	5	3.23	.69	H
2	8	2.49	.83	
2	11	2.14	.72	
3	2	2.86	.83	
3	5	2.91	.76	
3	8	2.58	.85	
3	11	2.51	.64	
4	2	2.70	.48	R
4	5	2.89	.55	R
4	8	3.09	.70	
4	11	3.18	.75	

Table 8.A.6 AIS and Polyserial Correlation: Level I, Mathematics

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/ Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	13	2.66	.72	
1	14	2.65	.73	
Operational	15	2.73	.76	
Operational	16	2.79	.73	
1	17	3.20	.70	
Operational	18	2.75	.74	
Operational	19	2.86	.73	
1	20	3.22	.71	
Operational	21	2.85	.77	
Operational	22	3.23	.76	
1	23	2.52	.71	
Operational	24	2.92	.78	
2	14	3.20	.72	
2	17	3.02	.65	
2	20	2.89	.64	
2	23	2.67	.65	
3	14	3.13	.72	
3	17	2.13	.61	
3	20	3.27	.73	
3	23	2.78	.62	
4	14	2.52	.74	
4	17	1.95	.65	
4	20	2.88	.69	
4	23	2.35	.74	

Table 8.A.7 AIS and Polyserial Correlation: Level II, Mathematics

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/ Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	13	3.25	.71	H
1	14	3.06	.75	
Operational	15	3.49	.78	H
Operational	16	2.12	.78	
1	17	2.72	.75	
Operational	18	1.38	.59	R
Operational	19	2.29	.62	
1	20	2.85	.65	
Operational	21	2.38	.84	
Operational	22	2.69	.77	
1	23	3.21	.67	H
Operational	24	2.79	.77	
2	14	3.18	.75	
2	17	3.47	.60	R H
2	20	2.86	.65	
2	23	2.93	.69	
3	14	3.37	.62	H
3	17	2.67	.71	
3	20	1.34	.53	R
3	23	2.84	.70	
4	14	3.10	.72	
4	17	2.84	.72	
4	20	3.40	.66	H
4	23	1.22	.64	

Table 8.A.8 AIS and Polyserial Correlation: Level III, Mathematics

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/ Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	13	2.17	.76	
1	14	2.98	.64	
Operational	15	2.04	.68	
Operational	16	2.98	.80	
1	17	2.09	.58	R
Operational	18	2.20	.66	
Operational	19	2.42	.62	
1	20	2.75	.70	
Operational	21	2.76	.65	
Operational	22	1.76	.63	
1	23	2.61	.45	R
Operational	24	2.25	.64	
2	14	2.38	.46	R
2	17	3.28	.62	H
2	20	3.23	.58	R H
2	23	2.88	.66	
3	14	2.13	.60	
3	17	3.52	.54	R H
3	20	3.04	.70	
3	23	2.14	.60	
4	14	2.72	.64	
4	17	2.81	.73	
4	20	2.13	.66	
4	23	2.34	.58	R

Table 8.A.9 AIS and Polyserial Correlation: Level IV, Mathematics

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/ Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	13	2.45	.84	
1	14	2.90	.60	
Operational	15	2.94	.85	
Operational	16	1.83	.79	
1	17	3.11	.64	
Operational	18	2.11	.67	
Operational	19	3.00	.62	
1	20	3.09	.65	
Operational	21	2.93	.85	
Operational	22	1.85	.78	
1	23	2.07	.53	R
Operational	24	2.62	.61	
2	14	1.48	.57	R
2	17	3.00	.84	
2	20	3.07	.59	R
2	23	3.04	.83	
3	14	3.39	.53	R H
3	17	2.93	.82	
3	20	1.90	.70	
3	23	2.58	.49	R
4	14	3.30	.65	H
4	17	2.21	.71	
4	20	1.85	.72	
4	23	2.92	.82	

Table 8.A.10 AIS and Polyserial Correlation: Level V, Mathematics

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/ Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	13	3.35	.81	H
1	14	3.38	.75	H
Operational	15	2.78	.74	
Operational	16	2.32	.80	
1	17	2.75	.75	
Operational	18	2.33	.77	
Operational	19	2.44	.72	
1	20	2.81	.72	
Operational	21	2.60	.72	
Operational	22	2.01	.76	
1	23	2.44	.70	
Operational	24	2.27	.77	
2	14	3.16	.71	
2	17	3.17	.75	
2	20	2.58	.67	
2	23	2.88	.72	
3	14	2.57	.78	
3	17	2.39	.69	
3	20	3.34	.82	H
3	23	3.16	.78	
4	14	3.34	.75	H
4	17	3.03	.65	
4	20	3.21	.76	H
4	23	3.26	.77	H

Table 8.A.11 AIS and Polyserial Correlation: Level I, Science

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	25	3.01	.81	
1/3*	26	2.88	.74	
Operational	27	2.66	.80	
Operational	28	2.92	.80	
1/3*	29	3.51	.72	
Operational	30	2.88	.83	
Operational	31	2.65	.82	
1/3*	32	2.87	.78	
Operational	33	2.83	.77	
Operational	34	3.50	.82	
1/3*	35	2.36	.73	
Operational	36	2.81	.76	
2/4*	26	3.38	.79	
2/4*	29	4.12	.57	R H
2/4*	32	3.10	.69	
2/4*	35	3.99	.67	

* This task appeared on more than one field-test form.

Table 8.A.12 AIS and Polyserial Correlation: Level III, Science

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	25	2.67	.61	
1/3*	26	2.71	.73	
Operational	27	2.94	.73	
Operational	28	2.65	.77	
1/3*	29	2.37	.50	R
Operational	30	2.50	.80	
Operational	31	2.73	.73	
1/3*	32	2.70	.71	
Operational	33	2.75	.76	
Operational	34	2.34	.67	
1/3*	35	2.63	.59	R
Operational	36	2.91	.74	
2/4*	26	2.74	.58	R
2/4*	29	2.58	.59	R
2/4*	32	2.20	.71	
2/4*	35	2.30	.58	R

* This task appeared on more than one field-test form.

Table 8.A.13 AIS and Polyserial Correlation: Level IV, Science

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/ Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	25	2.96	.75	
1/3*	26	1.97	.42	R
Operational	27	2.53	.73	
Operational	28	2.34	.70	
1/3*	29	3.44	.66	H
Operational	30	2.91	.67	
Operational	31	2.94	.70	
1/3*	32	2.62	.65	
Operational	33	2.58	.75	
Operational	34	2.42	.75	
1/3*	35	2.59	.64	
Operational	36	2.85	.75	
2/4*	26	2.13	.67	
2/4*	29	2.21	.56	R
2/4*	32	2.57	.64	
2/4*	35	2.86	.61	

* This task appeared on more than one field-test form.

Table 8.A.14 AIS and Polyserial Correlation: Level V, Science

Flag values are as follows:
A = low average task score
R = low correlation with criterion
O = high percent of omits/not responding
H = high average task score

Version/ Field-Test Form	Task Position	AIS	Polyserial	Flag
Operational	25	2.37	.73	
1/3*	26	2.28	.56	R
Operational	27	2.30	.69	
Operational	28	2.72	.77	
1/3*	29	2.21	.70	
Operational	30	2.14	.76	
Operational	31	2.43	.74	
1/3*	32	2.41	.67	
Operational	33	2.71	.77	
Operational	34	3.33	.81	H
1/3*	35	2.51	.70	
Operational	36	2.66	.77	
2/4*	26	1.80	.66	
2/4*	29	1.72	.55	R
2/4*	32	1.95	.62	
2/4*	35	2.21	.68	

* This task appeared on more than one field-test form.

Table 8.A.15 Frequency of Operational Task Scores: ELA

ELA Level	Score on Task	1		2		3		4		5		6		7		8	
		Count	Pct	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
I	0	1,121	7.94	1,669	11.82	1,458	10.33	1,748	12.38	1,502	10.64	1,783	12.63	1,733	12.28	1,871	13.25
	1	1,101	7.80	4,893	34.66	2,619	18.55	4,331	30.68	2,884	20.43	3,887	27.53	4,038	28.60	4,330	30.67
	2	228	1.61	776	5.50	566	4.01	827	5.86	603	4.27	1,155	8.18	644	4.56	657	4.65
	3	320	2.27	840	5.95	571	4.04	900	6.37	609	4.31	635	4.50	682	4.83	746	5.28
	4	1031	7.30	1,075	7.61	1,093	7.74	1,399	9.91	941	6.67	922	6.53	989	7.01	1,108	7.85
	5	10,317	73.08	4,865	34.46	7,811	55.33	4,913	34.80	7,579	53.68	5,736	40.63	6,032	42.73	5,406	38.29
II	0	264	3.76	329	4.69	245	3.49	280	3.99	543	7.74	279	3.98	364	5.19	469	6.68
	1	2,356	33.57	1,941	27.66	242	3.45	1,455	20.73	2,037	29.03	1,452	20.69	2,398	34.17	2,159	30.76
	2	626	8.92	1,330	18.95	491	7.00	1,624	23.14	1,204	17.16	2,104	29.98	2,656	37.85	1,817	25.89
	3	607	8.65	1,621	23.10	976	13.91	1,124	16.02	2,238	31.89	1,756	25.02	1,018	14.51	1,111	15.83
	4	3,165	45.10	1,797	25.61	5,064	72.16	2,535	36.12	996	14.19	1,427	20.33	582	8.29	1,462	20.83
III	0	228	3.04	498	6.63	356	4.74	338	4.50	587	7.82	305	4.06	635	8.45	390	5.19
	1	284	3.78	1,026	13.66	568	7.56	1,231	16.39	1,869	24.88	826	11.00	2,225	29.62	321	4.27
	2	440	5.86	1,840	24.50	1,207	16.07	3,205	42.67	1,146	15.26	907	12.08	1,356	18.05	442	5.88
	3	861	11.46	2,351	31.30	2,662	35.44	1,923	25.60	1,537	20.46	1,745	23.23	1,941	25.84	1,256	16.72
	4	5,698	75.86	1,796	23.91	2,718	36.19	814	10.84	2,372	31.58	3,728	49.63	1,354	18.03	5,102	67.93
IV	0	555	5.39	729	7.08	991	9.63	689	6.69	421	4.09	521	5.06	503	4.89	677	6.58
	1	2,705	26.27	2,125	20.64	5,670	55.08	2,172	21.10	452	4.39	2,006	19.49	1,770	17.19	4,268	41.46
	2	863	8.38	2,763	26.84	1,475	14.33	1,780	17.29	491	4.77	2,308	22.42	2,615	25.40	1,100	10.68
	3	1,317	12.79	2,064	20.05	976	9.48	2,751	26.72	2,361	22.93	2,483	24.12	2,801	27.21	2,113	20.52
	4	4,855	47.16	2,614	25.39	1,183	11.49	2,903	28.20	6,570	63.82	2,977	28.92	2,606	25.31	2,137	20.76
V	0	565	5.27	778	7.26	704	6.57	759	7.08	786	7.33	868	8.10	909	8.48	806	7.52
	1	563	5.25	1,974	18.42	1,366	12.75	4,802	44.81	1,132	10.56	2,588	24.15	2,758	25.73	448	4.18
	2	1,217	11.36	1,955	18.24	2,011	18.76	1,493	13.93	1,052	9.82	1,778	16.59	1,724	16.09	610	5.69
	3	4,506	42.05	3,685	34.38	3,536	32.99	2,222	20.73	1,953	18.22	3,098	28.91	2,674	24.95	1,860	17.36
	4	3,866	36.07	2,325	21.69	3,100	28.93	1,441	13.45	5,794	54.06	2,385	22.25	2,652	24.75	6,993	65.25

Table 8.A.16 Frequency of Operational Task Scores: Mathematics

Math Level	Score on Task	1		2		3		4		5		6		7		8	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
I	0	1,657	11.74	1,938	13.73	1,532	10.85	2,952	20.91	1,671	11.84	1,822	12.91	1,422	10.07	1,965	13.92
	1	5,065	35.88	4,529	32.08	5,020	35.56	3,573	25.31	4,474	31.69	4,405	31.20	3,760	26.63	4,036	28.59
	2	772	5.47	791	5.60	638	4.52	508	3.60	759	5.38	745	5.28	602	4.26	705	4.99
	3	920	6.52	828	5.86	791	5.60	659	4.67	784	5.55	782	5.54	726	5.14	802	5.68
	4	1,381	9.78	1,297	9.19	1,164	8.24	976	6.91	1,192	8.44	1,077	7.63	1,030	7.30	1,223	8.66
	5	4,323	30.62	4,735	33.54	4,973	35.22	5,450	38.60	5,238	37.10	5,287	37.45	6,578	46.59	5,387	38.16
II	0	236	3.36	260	3.70	394	5.61	572	8.15	335	4.77	429	6.11	432	6.16	359	5.12
	1	372	5.30	646	9.20	3,410	48.59	4,660	66.40	1,802	25.68	2,542	36.22	1,523	21.70	1,378	19.64
	2	710	10.12	453	6.45	601	8.56	959	13.66	2,499	35.61	814	11.60	1,304	18.58	1,167	16.63
	3	2,452	34.94	470	6.70	544	7.75	457	6.51	667	9.50	816	11.63	883	12.58	1,278	18.21
	4	3,248	46.28	5,189	73.94	2,069	29.48	370	5.27	1,715	24.44	2,417	34.44	2,876	40.98	2,836	40.41
III	0	349	4.65	392	5.22	332	4.42	378	5.03	297	3.95	308	4.10	373	4.97	401	5.34
	1	2,981	39.69	2,829	37.66	1,572	20.93	1,513	20.14	1,672	22.26	741	9.87	4,015	53.45	2,228	29.66
	2	1,053	14.02	1,698	22.61	751	10.00	3,158	42.05	1,903	25.34	1,281	17.05	1,475	19.64	1,886	25.11
	3	1,673	22.27	1,575	20.97	844	11.24	1,645	21.90	2,352	31.31	3,876	51.60	684	9.11	1,564	20.82
	4	1,455	19.37	1,017	13.54	4,012	53.41	817	10.88	1,287	17.13	1,305	17.37	964	12.83	1,432	19.07
IV	0	546	5.30	558	5.42	737	7.16	511	4.96	473	4.59	594	5.77	725	7.04	565	5.49
	1	4,052	39.36	2,325	22.58	5,958	57.87	4,234	41.13	1,079	10.48	2,295	22.29	4,388	42.62	1,650	16.03
	2	636	6.18	592	5.75	786	7.63	2,308	22.42	1,906	18.51	610	5.93	2,532	24.59	2,524	24.52
	3	1,091	10.60	1,533	14.89	650	6.31	868	8.43	2,597	25.23	1,557	15.12	1,739	16.89	3,061	29.73
	4	3,970	38.56	5,287	51.36	2,164	21.02	2,374	23.06	4,240	41.19	5,239	50.89	911	8.85	2,495	24.24
V	0	659	6.15	635	5.93	708	6.61	723	6.75	655	6.11	797	7.44	866	8.08	777	7.25
	1	1,586	14.80	1,096	10.23	4,304	40.16	3,312	30.90	2,941	27.44	3,684	34.38	4,451	41.53	3,696	34.49
	2	278	2.59	3,938	36.75	1,191	11.11	2,273	21.21	2,303	21.49	594	5.54	2,011	18.76	2,054	19.17
	3	605	5.65	748	6.98	1,064	9.93	1,717	16.02	1,860	17.36	883	8.24	1,513	14.12	1,569	14.64
	4	7,589	70.81	4,300	40.12	3,450	32.19	2,692	25.12	2,958	27.60	4,759	44.41	1,876	17.50	2,621	24.46

Table 8.A.17 Frequency of Operational Task Scores: Science

Science Level	Score on Task	1		2		3		4		5		6		7		8	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
I	0	816	18.66	847	19.37	826	18.89	800	18.29	832	19.03	821	18.77	737	16.85	837	19.14
	1	1,072	24.51	1,371	31.35	1,181	27.01	1,258	28.77	1,351	30.89	1,263	28.88	795	18.18	1,290	29.50
	2	234	5.35	242	5.53	216	4.94	229	5.24	280	6.40	217	4.96	171	3.91	231	5.28
	3	199	4.55	223	5.10	227	5.19	187	4.28	221	5.05	265	6.06	153	3.50	244	5.58
	4	325	7.43	296	6.77	338	7.73	262	5.99	286	6.54	300	6.86	290	6.63	314	7.18
	5	1,727	39.49	1,394	31.88	1,585	36.25	1,637	37.43	1,403	32.08	1,507	34.46	2,227	50.93	1,457	33.32
III	0	229	6.09	228	6.07	291	7.74	276	7.34	255	6.78	266	7.08	274	7.29	249	6.62
	1	446	11.86	230	6.12	712	18.94	1,163	30.94	491	13.06	773	20.56	855	22.75	439	11.68
	2	1,244	33.09	837	22.27	684	18.20	629	16.73	889	23.65	407	10.83	1,105	29.40	738	19.63
	3	749	19.93	1,309	34.82	951	25.30	357	9.50	1,027	27.32	1,080	28.73	852	22.67	947	25.19
	4	1,091	29.02	1,155	30.73	1,121	29.82	1,334	35.49	1,097	29.18	1,233	32.80	673	17.90	1,386	36.87
IV	0	245	6.84	248	6.92	269	7.51	288	8.04	245	6.84	273	7.62	283	7.90	276	7.70
	1	391	10.91	578	16.13	838	23.39	493	13.76	179	5.00	583	16.27	819	22.86	476	13.28
	2	790	22.05	1,066	29.75	1,075	30.00	588	16.41	676	18.87	750	20.93	816	22.77	579	16.16
	3	632	17.64	967	26.99	734	20.49	881	24.59	1,616	45.10	1,376	38.40	1,003	27.99	1,154	32.21
	4	1,525	42.56	724	20.21	667	18.62	1,333	37.20	867	24.20	601	16.77	662	18.48	1,098	30.64
V	0	395	9.98	463	11.69	426	10.76	468	11.82	424	10.71	431	10.89	417	10.53	463	11.69
	1	792	20.01	810	20.46	516	13.03	1,119	28.26	759	19.17	416	10.51	263	6.64	636	16.06
	2	976	24.65	1,131	28.57	666	16.82	1,106	27.94	1,063	26.85	676	17.08	392	9.90	695	17.55
	3	1,286	32.48	985	24.88	1,481	37.41	667	16.85	905	22.86	1,819	45.95	581	14.68	1,202	30.36
	4	510	12.88	570	14.40	870	21.98	599	15.13	808	20.41	617	15.58	2,306	58.25	963	24.32

Appendix 8.B—Reliability Analyses

Table 8.B.1 Reliabilities and SEMs by GENDER

Content Area	Level	No. of Tasks	Male		Female		Unknown Gender	
			Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
English–Language Arts	I	8	0.88	3.88	0.89	3.84	0.93	3.72
	II	8	0.83	2.59	0.81	2.57	0.68	2.77
	III	8	0.88	2.19	0.86	2.20	0.66	2.68
	IV	8	0.86	2.51	0.86	2.44	0.81	2.55
	V	8	0.89	2.20	0.90	2.18	0.88	2.18
Mathematics	I	8	0.84	4.28	0.85	4.26	0.90	3.93
	II	8	0.83	2.64	0.80	2.65	0.75	2.74
	III	8	0.78	2.63	0.74	2.63	0.71	2.69
	IV	8	0.86	2.64	0.84	2.65	0.84	2.44
	V	8	0.85	2.74	0.85	2.72	0.84	2.79
Science *	I	8	0.88	3.97	0.90	3.87	–	–
	III	8	0.85	2.66	0.83	2.67	–	–
	IV	8	0.85	2.39	0.83	2.34	–	–
	V	8	0.86	2.30	0.84	2.29	–	–

* Results for groups with fewer than 11 members are not reported.

Table 8.B.2 Reliabilities and SEMs by PRIMARY ETHNICITY

Content Area	Level	No. of Tasks	American Indian		Asian		Pacific Islander		Filipino	
			Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
English–Language Arts	I	8	0.90	3.58	0.86	3.97	0.89	3.76	0.87	3.97
	II	8	0.77	2.53	0.83	2.58	0.86	2.54	0.81	2.66
	III	8	0.91	2.06	0.90	2.27	0.86	2.22	0.89	2.32
	IV	8	0.82	2.48	0.84	2.46	0.88	2.35	0.86	2.47
	V	8	0.92	2.13	0.90	2.23	0.89	2.05	0.90	2.27
Mathematics	I	8	0.82	4.32	0.82	4.36	0.85	4.34	0.82	4.43
	II	8	0.78	2.58	0.84	2.64	0.84	2.63	0.83	2.63
	III	8	0.78	2.72	0.79	2.64	0.66	2.72	0.82	2.60
	IV	8	0.81	2.78	0.85	2.69	0.85	2.62	0.85	2.70
	V	8	0.85	2.76	0.87	2.68	0.85	2.65	0.85	2.73
Science	I	8	0.87	4.11	0.87	4.05	0.92	3.61	0.88	4.06
	III	8	0.85	2.36	0.87	2.67	0.76	2.42	0.86	2.41
	IV	8	0.84	2.23	0.81	2.37	0.81	3.05	0.74	2.38
	V	8	0.89	2.11	0.85	2.40	0.84	2.30	0.79	2.38

Content Area	Level	No. of Tasks	Hispanic		African American		White		Unknown Ethnicity	
			Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
English–Language Arts	I	8	0.89	3.80	0.89	3.76	0.87	3.96	0.84	4.19
	II	8	0.82	2.59	0.82	2.59	0.83	2.56	0.82	2.66
	III	8	0.87	2.17	0.87	2.13	0.88	2.20	0.86	2.31
	IV	8	0.86	2.45	0.86	2.45	0.86	2.51	0.85	2.60
	V	8	0.89	2.17	0.88	2.17	0.89	2.20	0.90	2.20
Mathematics	I	8	0.85	4.21	0.86	4.22	0.83	4.35	0.82	4.37
	II	8	0.82	2.64	0.82	2.64	0.82	2.64	0.81	2.74
	III	8	0.76	2.63	0.77	2.63	0.78	2.61	0.72	2.69
	IV	8	0.85	2.62	0.85	2.62	0.85	2.65	0.85	2.73
	V	8	0.85	2.70	0.84	2.75	0.84	2.77	0.85	2.79
Science	I	8	0.90	3.85	0.89	3.95	0.87	4.06	0.89	3.91
	III	8	0.83	2.66	0.85	2.66	0.84	2.75	0.87	2.38
	IV	8	0.84	2.33	0.86	2.31	0.84	2.45	0.86	2.57
	V	8	0.85	2.31	0.86	2.34	0.87	2.26	0.88	2.20

Table 8.B.3 Reliabilities and SEMs by PRIMARY ETHNICITY for Economically Disadvantaged

Content Area	Level	No. of Tasks	American Indian		Asian		Pacific Islander		Filipino	
			Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
English–Language Arts	I	8	0.88	3.45	0.86	3.95	0.90	3.66	0.87	3.97
	II	8	0.73	2.62	0.84	2.57	0.80	2.64	0.81	2.61
	III	8	0.90	2.04	0.89	2.25	0.88	2.30	0.89	2.34
	IV	8	0.84	2.38	0.85	2.40	0.86	2.38	0.86	2.43
	V	8	0.89	2.09	0.89	2.23	0.87	2.12	0.90	2.19
Mathematics	I	8	0.80	4.33	0.83	4.30	0.85	4.20	0.83	4.37
	II	8	0.77	2.59	0.84	2.61	0.85	2.58	0.84	2.57
	III	8	0.74	2.76	0.78	2.69	0.71	2.66	0.77	2.60
	IV	8	0.84	2.67	0.87	2.64	0.84	2.69	0.86	2.66
	V	8	0.85	2.68	0.86	2.68	0.79	2.78	0.87	2.60
Science *	I	8	0.88	4.02	0.86	4.07	0.92	3.51	0.86	4.13
	III	8	0.83	2.47	0.85	2.73	0.75	2.53	0.84	2.50
	IV	8	0.86	2.18	0.85	2.31	–	–	0.68	2.37
	V	8	0.85	2.16	0.81	2.34	–	–	0.72	2.28

Content Area	Level	No. of Tasks	Hispanic		African American		White		Unknown Ethnicity	
			Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
English–Language Arts	I	8	0.88	3.78	0.89	3.71	0.88	3.83	0.87	4.02
	II	8	0.81	2.59	0.81	2.58	0.83	2.55	0.82	2.65
	III	8	0.86	2.16	0.86	2.12	0.88	2.14	0.86	2.27
	IV	8	0.86	2.45	0.85	2.43	0.86	2.47	0.85	2.58
	V	8	0.89	2.16	0.87	2.13	0.89	2.16	0.91	2.16
Mathematics	I	8	0.85	4.21	0.85	4.19	0.83	4.30	0.86	4.13
	II	8	0.82	2.64	0.82	2.64	0.81	2.64	0.81	2.69
	III	8	0.75	2.64	0.76	2.62	0.77	2.61	0.70	2.67
	IV	8	0.85	2.62	0.85	2.61	0.86	2.63	0.85	2.70
	V	8	0.85	2.69	0.83	2.75	0.84	2.76	0.86	2.75
Science	I	8	0.90	3.84	0.90	3.88	0.87	3.99	0.89	3.71
	III	8	0.82	2.64	0.85	2.56	0.83	2.58	0.87	2.39
	IV	8	0.83	2.33	0.87	2.22	0.87	2.43	0.88	2.28
	V	8	0.85	2.31	0.85	2.26	0.87	2.27	0.88	2.23

* Results for groups with fewer than 11 members are not reported.

Table 8.B.4 Reliabilities and SEMs by PRIMARY ETHNICITY for Not Economically Disadvantaged

Content Area	Level	No. of Tasks	American Indian		Asian		Pacific Islander		Filipino	
			Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
English–Language Arts	I	8	0.89	3.77	0.86	4.00	0.89	3.88	0.87	3.95
	II	8	0.84	2.39	0.83	2.57	0.93	2.35	0.81	2.71
	III	8	0.92	2.08	0.90	2.29	0.81	2.05	0.89	2.27
	IV	8	0.81	2.65	0.83	2.48	0.91	2.28	0.85	2.48
	V	8	0.93	2.14	0.91	2.21	0.92	1.86	0.89	2.33
Mathematics	I	8	0.82	4.31	0.81	4.41	0.85	4.51	0.82	4.43
	II	8	0.80	2.56	0.84	2.65	0.81	2.83	0.82	2.67
	III	8	0.82	2.60	0.80	2.60	0.49	2.78	0.84	2.60
	IV	8	0.66	3.17	0.84	2.74	0.89	2.54	0.85	2.75
	V	8	0.87	2.79	0.88	2.68	0.91	2.37	0.84	2.79
Science *	I	8	0.86	4.09	0.88	4.05	–	–	0.89	4.05
	III	8	0.86	2.26	0.88	2.60	–	–	0.87	2.36
	IV	8	–	–	0.76	2.40	–	–	0.78	2.38
	V	8	0.93	2.05	0.86	2.49	–	–	0.82	2.43

Content Area	Level	No. of Tasks	Hispanic		African American		White		Unknown Ethnicity	
			Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
English–Language Arts	I	8	0.89	3.88	0.89	3.81	0.86	4.02	0.81	4.29
	II	8	0.84	2.58	0.83	2.62	0.83	2.57	0.84	2.62
	III	8	0.89	2.20	0.89	2.17	0.88	2.24	0.88	2.34
	IV	8	0.88	2.46	0.87	2.48	0.86	2.52	0.86	2.56
	V	8	0.90	2.21	0.89	2.24	0.90	2.22	0.87	2.34
Mathematics	I	8	0.86	4.21	0.86	4.23	0.82	4.37	0.77	4.51
	II	8	0.82	2.65	0.83	2.61	0.83	2.64	0.85	2.72
	III	8	0.78	2.59	0.77	2.65	0.77	2.60	0.74	2.61
	IV	8	0.86	2.64	0.86	2.61	0.84	2.66	0.84	2.73
	V	8	0.85	2.73	0.86	2.70	0.84	2.76	0.85	2.79
Science	I	8	0.90	3.86	0.89	4.00	0.87	4.10	0.90	3.98
	III	8	0.86	2.74	0.83	2.98	0.83	2.83	0.88	2.30
	IV	8	0.87	2.32	0.83	2.61	0.82	2.46	0.84	2.77
	V	8	0.86	2.29	0.86	2.49	0.86	2.26	0.90	2.21

* Results for groups with fewer than 11 members are not reported.

Table 8.B.5 Reliabilities and SEMs by PRIMARY ETHNICITY for Unknown Economic Status

Content Area	Level	No. of Tasks	American Indian		Asian		Pacific Islander		Filipino	
			Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
English–Language Arts *	I	8	–	–	0.90	3.82	–	–	0.67	4.25
	II	8	–	–	0.63	2.76	–	–	–	–
	III	8	–	–	0.93	2.14	–	–	–	–
	IV	8	–	–	0.73	2.45	–	–	–	–
	V	8	–	–	0.88	2.54	–	–	0.93	1.94
Mathematics *	I	8	–	–	0.82	4.40	–	–	0.40	5.07
	II	8	–	–	0.81	2.56	–	–	–	–
	III	8	–	–	0.76	2.74	–	–	–	–
	IV	8	–	–	0.74	2.70	–	–	–	–
	V	8	–	–	0.84	2.70	–	–	0.82	2.76
Science *	I	8	–	–	–	–	–	–	–	–
	III	8	–	–	–	–	–	–	–	–
	IV	8	–	–	–	–	–	–	–	–
	V	8	–	–	–	–	–	–	–	–

Content Area	Level	No. of Tasks	Hispanic		African American		White		Unknown Ethnicity	
			Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
English–Language Arts	I	8	0.89	3.74	0.86	4.24	0.88	3.99	0.86	4.20
	II	8	0.83	2.57	0.88	2.58	0.86	2.44	0.73	2.75
	III	8	0.86	2.32	0.71	2.14	0.85	2.05	0.69	2.22
	IV	8	0.89	2.42	0.74	2.64	0.86	2.56	0.87	2.69
	V	8	0.88	2.14	0.86	2.11	0.86	2.22	0.92	1.93
Mathematics	I	8	0.84	4.24	0.77	4.73	0.79	4.81	0.85	4.38
	II	8	0.86	2.55	0.78	2.53	0.86	2.61	0.72	2.80
	III	8	0.77	2.88	0.44	2.48	0.78	2.64	0.65	3.11
	IV	8	0.87	2.58	0.71	3.01	0.80	2.87	0.85	2.80
	V	8	0.81	2.82	0.72	3.01	0.79	2.95	0.81	2.90
Science *	I	8	0.85	4.40	–	–	0.91	3.78	–	–
	III	8	0.85	2.43	–	–	0.86	3.33	–	–
	IV	8	0.58	2.26	–	–	–	–	0.81	3.24
	V	8	0.68	2.47	–	–	0.91	1.97	0.83	2.03

* Results for groups with fewer than 11 members are not reported.

Table 8.B.6 Reliabilities and SEMs by Disability

Content Area	Level	No. of Tasks	Mental Retardation		Hard of Hearing		Deafness		Speech Impairment	
			Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
English–Language Arts	I	8	0.87	3.74	0.84	3.95	0.84	3.92	0.90	3.04
	II	8	0.79	2.57	0.82	2.44	0.85	2.62	0.71	2.63
	III	8	0.86	2.19	0.87	2.16	0.83	2.31	0.73	2.04
	IV	8	0.86	2.45	0.83	2.50	0.84	2.43	0.74	2.34
	V	8	0.89	2.16	0.90	2.12	0.86	2.14	0.77	2.06
Mathematics	I	8	0.82	4.31	0.87	4.01	0.80	4.31	0.82	3.86
	II	8	0.80	2.64	0.76	2.68	0.85	2.59	0.72	2.60
	III	8	0.75	2.56	0.75	2.70	0.81	2.55	0.62	2.71
	IV	8	0.84	2.62	0.85	2.68	0.83	2.65	0.73	2.61
	V	8	0.84	2.70	0.79	2.85	0.82	2.64	0.74	2.82
Science *	I	8	0.88	3.86	0.89	3.51	0.68	5.95	–	–
	III	8	0.82	2.67	0.86	2.38	0.86	2.31	0.72	3.19
	IV	8	0.82	2.36	0.88	2.10	0.69	2.29	0.76	2.23
	V	8	0.83	2.29	0.86	2.25	0.76	2.14	0.72	2.30

Content Area	Level	No. of Tasks	Visual Impairment		Emotional Disturbance		Orthopedic Impairment		Other Health Impairment	
			Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
English–Language Arts *	I	8	0.90	3.77	–	–	0.88	3.94	0.90	3.72
	II	8	0.80	2.47	0.73	2.73	0.82	2.55	0.81	2.56
	III	8	0.93	2.14	0.80	1.95	0.87	2.21	0.85	2.11
	IV	8	0.89	2.39	0.83	2.45	0.87	2.50	0.84	2.42
	V	8	0.87	2.23	0.85	2.03	0.90	2.15	0.87	2.04
Mathematics *	I	8	0.87	4.03	–	–	0.87	4.08	0.88	4.03
	II	8	0.81	2.71	0.75	2.30	0.82	2.63	0.82	2.59
	III	8	0.83	2.69	0.69	2.66	0.77	2.63	0.71	2.61
	IV	8	0.86	2.62	0.82	2.67	0.87	2.60	0.83	2.66
	V	8	0.84	2.74	0.80	2.79	0.86	2.69	0.81	2.76
Science *	I	8	0.91	3.81	–	–	0.90	3.76	0.91	3.78
	III	8	0.84	2.59	0.81	2.28	0.80	2.76	0.79	2.82
	IV	8	0.93	2.11	0.58	2.79	0.87	2.32	0.84	2.38
	V	8	0.86	2.06	0.82	2.23	0.90	2.26	0.86	2.17

Content Area	Level	No. of Tasks	Specific Learning Disability		Deaf-Blindness		Multiple Disabilities		Autism	
			Reliab.	SEM	Reliab.	SEM	Reliab.	SEM	Reliab.	SEM
English–Language Arts *	I	8	0.94	2.22	0.85	4.02	0.89	3.89	0.82	3.97
	II	8	0.67	2.52	–	–	0.81	2.66	0.85	2.56
	III	8	0.73	1.81	–	–	0.86	2.29	0.90	2.24
	IV	8	0.72	2.28	–	–	0.86	2.50	0.87	2.43
	V	8	0.73	1.91	–	–	0.90	2.31	0.90	2.26
Mathematics *	I	8	0.85	3.33	0.87	3.88	0.87	4.07	0.77	4.46
	II	8	0.69	2.53	–	–	0.83	2.60	0.84	2.63
	III	8	0.63	2.63	–	–	0.75	2.55	0.80	2.63
	IV	8	0.74	2.51	–	–	0.83	2.69	0.86	2.69
	V	8	0.70	2.73	–	–	0.87	2.64	0.86	2.76
Science *	I	8	0.86	2.93	0.90	3.48	0.89	3.90	0.81	4.29
	III	8	0.67	2.96	–	–	0.80	2.72	0.86	2.64
	IV	8	0.72	2.17	–	–	0.84	2.20	0.85	2.48
	V	8	0.72	2.51	–	–	0.90	2.23	0.87	2.38

Content Area	Level	No. of Tasks	Traumatic Brain Injury		Unknown Disability	
			Reliab.	SEM	Reliab.	SEM
English–Language Arts	I	8	0.93	3.52	0.86	4.12
	II	8	0.78	2.63	0.79	2.61
	III	8	0.83	2.16	0.82	2.15
	IV	8	0.76	2.57	0.84	2.55
	V	8	0.86	2.18	0.90	2.11
Mathematics	I	8	0.93	3.64	0.82	4.48
	II	8	0.86	2.50	0.78	2.73
	III	8	0.73	2.57	0.70	2.73
	IV	8	0.85	2.52	0.83	2.67
	V	8	0.84	2.83	0.81	2.84
Science	I	8	0.97	2.97	0.88	4.32
	III	8	0.71	2.71	0.81	2.79
	IV	8	0.70	2.22	0.83	2.44
	V	8	0.84	2.36	0.85	2.19

* Results for groups with fewer than 11 members are not reported.

Table 8.B.7 Decision Accuracy and Decision Consistency: Level I, ELA

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy All-forms Average *	26–40	0.50	0.05	0.00	0.00	0.00	0.56
	16–25	0.05	0.15	0.03	0.00	0.00	0.24
	10–15	0.00	0.03	0.05	0.01	0.00	0.09
	5–9	0.00	0.00	0.02	0.03	0.01	0.06
	0–4	0.00	0.00	0.01	0.02	0.02	0.05
	Estimated Proportion Correctly Classified: Total = 0.75, Proficient & Above = 0.93						
Decision Consistency Alternate Form *	26–40	0.49	0.07	0.00	0.00	0.00	0.56
	16–25	0.07	0.12	0.04	0.01	0.00	0.24
	10–15	0.00	0.03	0.03	0.02	0.00	0.09
	5–9	0.00	0.01	0.02	0.02	0.01	0.06
	0–4	0.00	0.00	0.01	0.01	0.02	0.05
	Estimated Proportion Correctly Classified: Total = 0.68 , Proficient & Above = 0.90						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Table 8.B.8 Decision Accuracy and Decision Consistency: Level I, Mathematics

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy	29–40	0.28	0.06	0.00	0.00	0.00	0.33
	18–28	0.06	0.24	0.05	0.01	0.00	0.36
	10–17	0.00	0.05	0.10	0.02	0.00	0.16
All-forms Average *	5–9	0.00	0.00	0.04	0.04	0.01	0.08
	0–4	0.00	0.00	0.01	0.03	0.03	0.07
	Estimated Proportion Correctly Classified: Total = 0.67, Proficient & Above = 0.88						
Decision Consistency	29–40	0.26	0.07	0.00	0.00	0.00	0.33
	18–28	0.09	0.18	0.07	0.01	0.00	0.36
	10–17	0.00	0.06	0.08	0.03	0.01	0.16
Alternate Form *	5–9	0.00	0.01	0.03	0.03	0.02	0.08
	0–4	0.00	0.00	0.01	0.02	0.03	0.07
	Estimated Proportion Correctly Classified: Total = 0.58, Proficient & Above = 0.84						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Table 8.B.9 Decision Accuracy and Decision Consistency: Level I, Science

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy	30–40	0.31	0.05	0.00	0.00	0.00	0.35
	21–29	0.04	0.15	0.04	0.00	0.00	0.23
	12–20	0.01	0.05	0.13	0.03	0.00	0.22
All-forms Average *	5–11	0.00	0.00	0.04	0.07	0.01	0.12
	0–4	0.00	0.00	0.00	0.03	0.04	0.07
	Estimated Proportion Correctly Classified: Total = 0.70, Proficient & Above = 0.90						
Decision Consistency	30–40	0.29	0.06	0.00	0.00	0.00	0.35
	21–29	0.06	0.11	0.05	0.00	0.00	0.23
	12–20	0.01	0.06	0.10	0.04	0.01	0.22
Alternate Form *	5–11	0.00	0.00	0.04	0.06	0.02	0.12
	0–4	0.00	0.00	0.00	0.02	0.04	0.07
	Estimated Proportion Correctly Classified: Total = 0.61, Proficient & Above = 0.87						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Table 8.B.10 Decision Accuracy and Decision Consistency: Level II, ELA

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy	22–32	0.36	0.07	0.00	0.00	0.00	0.42
	14–21	0.07	0.32	0.03	0.00	0.00	0.43
	10–13	0.00	0.04	0.05	0.01	0.00	0.10
All-forms Average *	4–9	0.00	0.00	0.02	0.02	0.00	0.04
	0–3	0.00	0.00	0.00	0.01	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.75, Proficient & Above = 0.92						
Decision Consistency	22–32	0.34	0.09	0.00	0.00	0.00	0.42
	14–21	0.10	0.27	0.05	0.01	0.00	0.43
	10–13	0.00	0.05	0.04	0.02	0.00	0.10
Alternate Form *	4–9	0.00	0.00	0.01	0.02	0.00	0.04
	0–3	0.00	0.00	0.00	0.00	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.66, Proficient & Above = 0.89						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Table 8.B.11 Decision Accuracy and Decision Consistency: Level II, Mathematics

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy	24–32	0.27	0.07	0.00	0.00	0.00	0.35
	18–23	0.05	0.22	0.05	0.00	0.00	0.31
	13–17	0.01	0.07	0.12	0.03	0.00	0.23
All–forms Average *	7–12	0.00	0.00	0.03	0.05	0.00	0.09
	0–6	0.00	0.00	0.00	0.01	0.00	0.02
	Estimated Proportion Correctly Classified: Total = 0.68, Proficient & Above = 0.87						
Decision Consistency	24–32	0.26	0.08	0.01	0.00	0.00	0.35
	18–23	0.07	0.17	0.06	0.01	0.00	0.31
	13–17	0.01	0.08	0.09	0.04	0.00	0.23
Alternate Form *	7–12	0.00	0.00	0.03	0.05	0.01	0.09
	0–6	0.00	0.00	0.00	0.01	0.01	0.02
	Estimated Proportion Correctly Classified: Total = 0.58, Proficient & Above = 0.83						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Table 8.B.12 Decision Accuracy and Decision Consistency: Level III, ELA

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy	25–32	0.40	0.06	0.00	0.00	0.00	0.46
	17–24	0.05	0.30	0.03	0.00	0.00	0.38
	10–16	0.00	0.03	0.09	0.00	0.00	0.12
All-forms Average *	4–9	0.00	0.00	0.01	0.02	0.00	0.03
	0–3	0.00	0.00	0.00	0.00	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.80, Proficient & Above = 0.93						
Decision Consistency	25–32	0.38	0.08	0.00	0.00	0.00	0.46
	17–24	0.08	0.25	0.05	0.00	0.00	0.38
	10–16	0.00	0.03	0.07	0.01	0.00	0.12
Alternate Form *	4–9	0.00	0.00	0.01	0.02	0.00	0.03
	0–3	0.00	0.00	0.00	0.00	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.72, Proficient & Above = 0.91						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Table 8.B.13 Decision Accuracy and Decision Consistency: Level III, Mathematics

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy	24–32	0.12	0.07	0.00	0.00	0.00	0.18
	16–23	0.05	0.43	0.06	0.00	0.00	0.54
	11–15	0.00	0.07	0.12	0.02	0.00	0.21
All–forms Average *	4–10	0.00	0.00	0.02	0.03	0.00	0.05
	0–3	0.00	0.00	0.00	0.01	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.70, Proficient & Above = 0.86						
Decision Consistency	24–32	0.11	0.07	0.00	0.00	0.00	0.18
	16–23	0.09	0.36	0.09	0.01	0.00	0.54
	11–15	0.00	0.08	0.10	0.03	0.00	0.21
Alternate Form *	4–10	0.00	0.00	0.02	0.03	0.00	0.05
	0–3	0.00	0.00	0.00	0.00	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.60, Proficient & Above = 0.82						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Table 8.B.14 Decision Accuracy and Decision Consistency: Level III, Science

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy	28–32	0.12	0.06	0.00	0.00	0.00	0.18
	19–27	0.03	0.43	0.05	0.00	0.00	0.50
	11–18	0.00	0.06	0.18	0.01	0.00	0.26
All-forms Average *	5–10	0.00	0.00	0.02	0.02	0.00	0.04
	0–4	0.00	0.00	0.00	0.01	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.76, Proficient & Above = 0.89						
Decision Consistency	28–32	0.12	0.07	0.00	0.00	0.00	0.18
	19–27	0.06	0.37	0.07	0.00	0.00	0.50
	11–18	0.00	0.08	0.15	0.03	0.00	0.26
Alternate Form *	5–10	0.00	0.00	0.02	0.02	0.00	0.04
	0–4	0.00	0.00	0.00	0.00	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.66, Proficient & Above = 0.85						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Table 8.B.15 Decision Accuracy and Decision Consistency: Level IV, ELA

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy	24–32	0.29	0.06	0.00	0.00	0.00	0.35
	15–23	0.05	0.35	0.04	0.00	0.00	0.44
	11–14	0.00	0.05	0.07	0.02	0.00	0.14
All-forms Average *	5–10	0.00	0.00	0.02	0.04	0.00	0.06
	0–4	0.00	0.00	0.00	0.01	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.75, Proficient & Above = 0.91						
Decision Consistency	24–32	0.28	0.07	0.00	0.00	0.00	0.35
	15–23	0.08	0.30	0.05	0.01	0.00	0.44
	11–14	0.00	0.06	0.05	0.03	0.00	0.14
Alternate Form *	5–10	0.00	0.00	0.02	0.04	0.00	0.06
	0–4	0.00	0.00	0.00	0.01	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.67, Proficient & Above = 0.88						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Table 8.B.16 Decision Accuracy and Decision Consistency: Level IV, Mathematics

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy	24–32	0.26	0.06	0.00	0.00	0.00	0.32
	17–23	0.05	0.25	0.05	0.00	0.00	0.35
	12–16	0.00	0.05	0.11	0.02	0.00	0.18
All-forms Average *	7–11	0.00	0.01	0.04	0.07	0.01	0.13
	0–6	0.00	0.00	0.00	0.02	0.01	0.02
	Estimated Proportion Correctly Classified: Total = 0.69, Proficient & Above = 0.89						
Decision Consistency	24–32	0.25	0.07	0.00	0.00	0.00	0.32
	17–23	0.07	0.20	0.07	0.01	0.00	0.35
	12–16	0.00	0.06	0.08	0.04	0.00	0.18
Alternate Form *	7–11	0.00	0.01	0.04	0.05	0.02	0.13
	0–6	0.00	0.00	0.00	0.01	0.01	0.02
	Estimated Proportion Correctly Classified: Total = 0.59, Proficient & Above = 0.85						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Table 8.B.17 Decision Accuracy and Decision Consistency: Level IV, Science

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy	27–32	0.14	0.06	0.00	0.00	0.00	0.20
	20–26	0.04	0.35	0.05	0.00	0.00	0.45
	13–19	0.00	0.07	0.19	0.02	0.00	0.29
All-forms Average *	5–12	0.00	0.00	0.02	0.04	0.00	0.06
	0–4	0.00	0.00	0.00	0.01	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.73, Proficient & Above = 0.88						
Decision Consistency	27–32	0.13	0.07	0.00	0.00	0.00	0.20
	20–26	0.07	0.29	0.08	0.00	0.00	0.45
	13–19	0.01	0.08	0.16	0.04	0.00	0.29
Alternate Form *	5–12	0.00	0.00	0.02	0.04	0.00	0.06
	0–4	0.00	0.00	0.00	0.01	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.63 , Proficient & Above = 0.84						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Table 8.B.18 Decision Accuracy and Decision Consistency: Level V, ELA

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy	24–32	0.42	0.06	0.00	0.00	0.00	0.48
	17–23	0.04	0.23	0.03	0.00	0.00	0.31
	10–16	0.00	0.04	0.12	0.01	0.00	0.17
All-forms Average *	4–9	0.00	0.00	0.01	0.02	0.00	0.04
	0–3	0.00	0.00	0.00	0.01	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.79, Proficient & Above = 0.93						
Decision Consistency	24–32	0.40	0.07	0.00	0.00	0.00	0.48
	17–23	0.07	0.19	0.05	0.00	0.00	0.31
	10–16	0.00	0.05	0.10	0.02	0.00	0.17
Alternate Form *	4–9	0.00	0.00	0.01	0.02	0.00	0.04
	0–3	0.00	0.00	0.00	0.00	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.72, Proficient & Above = 0.90						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Table 8.B.19 Decision Accuracy and Decision Consistency: Level V, Mathematics

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy	24–32	0.29	0.06	0.00	0.00	0.00	0.34
	17–23	0.05	0.23	0.05	0.00	0.00	0.34
	11–16	0.00	0.06	0.14	0.02	0.00	0.22
All-forms Average *	6–10	0.00	0.00	0.03	0.05	0.00	0.08
	0–5	0.00	0.00	0.00	0.02	0.00	0.02
	Estimated Proportion Correctly Classified: Total = 0.70, Proficient & Above = 0.89						
Decision Consistency	24–32	0.27	0.07	0.00	0.00	0.00	0.34
	17–23	0.08	0.18	0.07	0.01	0.00	0.34
	11–16	0.01	0.07	0.11	0.04	0.00	0.22
Alternate Form *	6–10	0.00	0.00	0.03	0.04	0.01	0.08
	0–5	0.00	0.00	0.00	0.01	0.00	0.02
	Estimated Proportion Correctly Classified: Total = 0.61, Proficient & Above = 0.85						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Table 8.B.20 Decision Accuracy and Decision Consistency: Level V, Science

	Placement Score	Advanced	Proficient	Basic	Below Basic	Far Below Basic	Category Total †
Decision Accuracy	26–32	0.16	0.04	0.00	0.00	0.00	0.20
	19–25	0.05	0.35	0.06	0.01	0.00	0.48
	11–18	0.00	0.05	0.21	0.00	0.00	0.25
All-forms Average *	3–10	0.00	0.00	0.02	0.03	0.00	0.05
	0–2	0.00	0.00	0.00	0.00	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.75, Proficient & Above = 0.88						
Decision Consistency	26–32	0.15	0.05	0.00	0.00	0.00	0.20
	19–25	0.08	0.30	0.09	0.01	0.00	0.48
	11–18	0.00	0.06	0.18	0.01	0.00	0.25
Alternate Form *	3–10	0.00	0.00	0.02	0.03	0.00	0.05
	0–2	0.00	0.00	0.00	0.00	0.00	0.01
	Estimated Proportion Correctly Classified: Total = 0.66, Proficient & Above = 0.84						

* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

Appendix 8.C—Correlations Between Content Area Scores

Note: Results for groups with fewer than 11 members are not reported.

Table 8.C.1 CAPA Content Area Correlations by Gender: Level I

	Male			Female			Unknown		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**8,719**	0.78	0.81	**4,967**	0.80	0.83	**33**	0.85	**–**
Mathematics	8,698	**8,701**	0.78	4,950	**4,956**	0.77	32	**32**	**–**
Science	2,207	2,208	**2,209**	1,295	1,295	**1,296**	7	7	**7**

Table 8.C.2 CAPA Content Area Correlations by Gender: Level II

	Male			Female			Unknown		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**4,573**	0.75	N/A	**2,033**	0.71	N/A	**37**	0.69	N/A
Mathematics	4,560	**4,561**	N/A	2,025	**2,026**	N/A	37	**37**	N/A
Science	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 8.C.3 CAPA Content Area Correlations by Gender: Level III

	Male			Female			Unknown		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**4,743**	0.73	0.69	**2,349**	0.70	0.70	**20**	0.66	**–**
Mathematics	4,729	**4,730**	0.68	2,346	**2,348**	0.69	20	**20**	**–**
Science	2,203	2,203	**2,204**	1,179	1,178	**1,180**	7	7	**7**

Table 8.C.4 CAPA Content Area Correlations by Gender: Level IV

	Male			Female			Unknown		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**6,487**	0.77	0.73	**3,344**	0.76	0.70	**27**	0.59	**–**
Mathematics	6,471	**6,472**	0.72	3,340	**3,345**	0.70	27	**28**	**–**
Science	2,049	2,047	**2,049**	1,099	1,098	**1,100**	6	6	**6**

Table 8.C.5 CAPA Content Area Correlations by Gender: Level V

	Male			Female			Unknown		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**6,500**	0.70	0.71	**3,668**	0.71	0.68	**49**	0.62	**–**
Mathematics	6,476	**6,486**	0.71	3,655	**3,661**	0.67	49	**49**	**–**
Science	2,031	2,032	**2,034**	1,201	1,201	**1,204**	7	7	**7**

Table 8.C.6 CAPA Content Area Correlations by Ethnicity: Level I

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**97**	0.77	0.77	**1,028**	0.80	0.83	**76**	0.85	0.89	**412**	0.75	0.76
Mathematics	97	**97**	0.81	1,023	**1,024**	0.73	76	**76**	0.87	410	**410**	0.76
Science	25	25	**25**	273	273	**273**	22	22	**22**	101	101	**101**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**7,256**	0.80	0.82	**1,268**	0.79	0.83	**3,197**	0.77	0.83	**385**	0.77	0.79
Mathematics	7,236	**7,241**	0.78	1,265	**1,266**	0.78	3,189	**3,191**	0.77	384	**384**	0.70
Science	1,889	1,890	**1,891**	335	335	**335**	778	778	**779**	86	86	**86**

Table 8.C.7 CAPA Content Area Correlations by Ethnicity: Level II

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**59**	0.75	N/A	**440**	0.73	N/A	**40**	0.68	N/A	**182**	0.75	N/A
Mathematics	59	**59**	N/A	438	**438**	N/A	40	**40**	N/A	181	**181**	N/A
Science	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**3,571**	0.74	N/A	**596**	0.73	N/A	**1,492**	0.75	N/A	**263**	0.68	N/A
Mathematics	3,563	**3,563**	N/A	594	**594**	N/A	1,485	**1,487**	N/A	262	**262**	N/A
Science	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 8.C.8 CAPA Content Area Correlations by Ethnicity: Level III

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**61**	0.69	0.53	**435**	0.72	0.72	**36**	0.70	0.71	**204**	0.78	0.86
Mathematics	61	**61**	0.58	434	**434**	0.74	36	**36**	0.57	204	**204**	0.75
Science	23	23	**23**	197	197	**197**	16	16	**16**	101	101	**101**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**3,838**	0.72	0.68	**681**	0.69	0.72	**1,642**	0.75	0.68	**215**	0.65	0.63
Mathematics	3,830	**3,832**	0.67	679	**679**	0.66	1,638	**1,639**	0.67	213	**213**	0.74
Science	1,823	1,824	**1,824**	320	319	**320**	818	817	**819**	91	91	**91**

Table 8.C.9 CAPA Content Area Correlations by Ethnicity: Level IV

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**80**	0.72	0.65	**619**	0.75	0.74	**51**	0.84	0.62	**318**	0.79	0.50
Mathematics	80	**80**	0.73	615	**616**	0.77	51	**51**	0.41	317	**317**	0.53
Science	26	26	**26**	182	181	**182**	13	13	**13**	79	79	**79**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**5,015**	0.77	0.72	**1,037**	0.76	0.73	**2,415**	0.76	0.71	**323**	0.75	0.80
Mathematics	5,004	**5,007**	0.71	1,036	**1,036**	0.71	2,412	**2,414**	0.73	323	**324**	0.75
Science	1,659	1,658	**1,659**	329	328	**329**	776	776	**777**	90	90	**90**

Table 8.C.10 CAPA Content Area Correlations by Ethnicity: Level V

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**107**	0.79	0.67	**619**	0.69	0.75	**48**	0.53	0.65	**292**	0.74	0.79
Mathematics	107	**107**	0.78	617	**617**	0.66	48	**48**	0.76	291	**292**	0.72
Science	34	34	**34**	193	193	**193**	14	14	**14**	85	85	**85**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**4,870**	0.70	0.67	**1,103**	0.68	0.70	**2,846**	0.71	0.69	**332**	0.69	0.81
Mathematics	4,856	**4,862**	0.68	1,094	**1,094**	0.71	2,835	**2,843**	0.72	332	**333**	0.78
Science	1,501	1,502	**1,505**	362	360	**362**	933	935	**935**	117	117	**117**

Table 8.C.11 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level I

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**59**	0.72	0.80	**459**	0.81	0.82	**43**	0.85	0.93	**140**	0.79	0.74
Mathematics	59	**59**	0.90	456	**457**	0.74	43	**43**	0.90	140	**140**	0.76
Science	14	14	**14**	135	135	**135**	14	14	**14**	37	37	**37**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**5,397**	0.79	0.82	**812**	0.78	0.82	**1,080**	0.78	0.85	**145**	0.77	0.74
Mathematics	5,379	**5,382**	0.78	811	**812**	0.75	1,078	**1,078**	0.75	144	**144**	0.62
Science	1,425	1,425	**1,425**	203	203	**203**	264	264	**264**	32	32	**32**

Table 8.C.12 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level II

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**44**	0.81	N/A	**193**	0.71	N/A	**24**	0.62	N/A	**60**	0.74	N/A
Mathematics	44	**44**	N/A	193	**193**	N/A	24	**24**	N/A	60	**60**	N/A
Science	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**2,964**	0.74	N/A	**441**	0.72	N/A	**627**	0.73	N/A	**119**	0.72	N/A
Mathematics	2,958	**2,958**	N/A	439	**439**	N/A	623	**623**	N/A	118	**118**	N/A
Science	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 8.C.13 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level III

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**43**	0.70	0.71	**178**	0.72	0.75	**25**	0.70	0.71	**75**	0.86	0.89
Mathematics	43	**43**	0.60	178	**178**	0.77	25	**25**	0.48	75	**75**	0.84
Science	12	12	**12**	84	84	**84**	12	12	**12**	35	35	**35**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**3,215**	0.71	0.65	**500**	0.66	0.71	**675**	0.72	0.64	**103**	0.61	0.55
Mathematics	3,209	**3,210**	0.67	498	**498**	0.68	674	**674**	0.59	103	**103**	0.79
Science	1,525	1,526	**1,526**	237	236	**237**	338	338	**338**	41	41	**41**

Table 8.C.14 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level IV

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**56**	0.74	0.66	**277**	0.74	0.78	**36**	0.87	**–**	**124**	0.73	0.34
Mathematics	56	**56**	0.77	277	**277**	0.82	36	**36**	**–**	124	**124**	0.50
Science	20	20	**20**	88	88	**88**	9	9	**9**	39	39	**39**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**4,177**	0.77	0.70	**774**	0.74	0.72	**983**	0.75	0.73	**157**	0.74	0.74
Mathematics	4,167	**4,169**	0.70	773	**773**	0.72	981	**982**	0.77	157	**157**	0.72
Science	1,398	1,397	**1,398**	254	253	**254**	300	300	**301**	49	49	**49**

Table 8.C.15 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level V

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**58**	0.77	0.71	**278**	0.68	0.77	**26**	0.37	**–**	**94**	0.66	0.73
Mathematics	58	**58**	0.86	276	**276**	059	26	**26**	**–**	94	**94**	0.58
Science	20	20	**20**	91	91	**91**	7	7	**7**	27	27	**27**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**3,836**	0.69	0.67	**747**	0.66	0.69	**1,087**	0.70	0.65	**162**	0.73	0.82
Mathematics	3,826	**3,830**	0.68	742	**742**	0.70	1,081	**1,083**	0.70	162	**163**	0.81
Science	1,218	1,218	**1,221**	252	251	**252**	388	389	**389**	56	56	**56**

Table 8.C.16 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level I

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**38**	0.84	0.86	**547**	0.79	0.82	**32**	0.83	**–**	**260**	0.73	0.77
Mathematics	38	**38**	0.76	545	**545**	0.70	32	**32**	**–**	258	**258**	0.77
Science	11	11	**11**	132	132	**132**	8	8	**8**	62	62	**62**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**1,741**	0.82	0.84	**429**	0.82	0.85	**2,048**	0.77	0.81	**204**	0.77	0.79
Mathematics	1,739	**1,741**	0.80	427	**427**	0.83	2,042	**2,044**	0.77	204	**204**	0.69
Science	439	440	**441**	125	125	**125**	501	501	**502**	45	45	**45**

Table 8.C.17 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level II

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**13**	0.85	N/A	**233**	0.76	N/A	**14**	0.79	N/A	**117**	0.77	N/A
Mathematics	13	**13**	N/A	231	**231**	N/A	14	**14**	N/A	116	**116**	N/A
Science	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**537**	0.74	N/A	**138**	0.77	N/A	**810**	0.76	N/A	**101**	0.67	N/A
Mathematics	535	**535**	N/A	138	**138**	N/A	807	**809**	N/A	101	**101**	N/A
Science	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 8.C.18 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level III

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**18**	0.71	0.38	**246**	0.73	0.72	**11**	0.78	**–**	**123**	0.76	0.85
Mathematics	18	**18**	0.74	245	**245**	0.72	11	**11**	**–**	123	**123**	0.71
Science	11	11	**11**	111	111	**111**	4	4	**4**	63	63	**63**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**566**	0.76	0.79	**170**	0.76	0.68	**922**	0.76	0.70	**89**	0.69	0.76
Mathematics	564	**565**	0.70	170	**170**	0.65	920	**921**	0.72	87	**87**	0.74
Science	276	276	**276**	77	77	**77**	462	462	**463**	41	41	**41**

Table 8.C.19 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level IV

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**22**	0.69	**–**	**324**	0.78	0.68	**15**	0.80	**–**	**187**	0.82	0.66
Mathematics	22	**22**	**–**	321	**322**	0.71	15	**15**	**–**	186	**186**	0.59
Science	5	5	**5**	89	88	**89**	4	4	**4**	40	40	**40**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**771**	0.77	0.79	**244**	0.80	0.78	**1,383**	0.76	0.68	**113**	0.80	0.83
Mathematics	770	**771**	0.79	244	**244**	0.68	1,382	**1,383**	0.70	113	**113**	0.78
Science	242	242	**242**	68	68	**68**	466	466	**466**	27	27	**27**

Table 8.C.20 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level V

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**46**	0.82	0.85	**326**	0.70	0.71	**20**	0.72	**–**	**186**	0.77	0.84
Mathematics	46	**46**	0.91	326	**326**	0.71	20	**20**	**–**	185	**186**	0.77
Science	14	14	**14**	96	96	**96**	6	6	**6**	55	55	**55**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**897**	0.71	0.70	**313**	0.72	0.77	**1,667**	0.72	0.70	**117**	0.66	0.80
Mathematics	893	**895**	0.65	310	**310**	0.72	1,662	**1,668**	0.72	117	**117**	0.81
Science	247	248	**248**	104	103	**104**	523	524	**524**	45	45	**45**

Table 8.C.21 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level I

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	0	–	–	**22**	0.84	**–**	**1**	–	–	**12**	0.35	–
Mathematics	0	0	–	22	**22**	**–**	1	**1**	–	12	**12**	–
Science	0	0	0	6	6	**6**	0	0	**0**	2	2	**2**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**118**	0.70	0.67	**27**	0.81	**–**	**69**	0.70	0.83	**36**	0.81	**–**
Mathematics	118	**118**	0.44	27	**27**	**–**	69	**69**	0.72	36	**36**	**–**
Science	25	25	**25**	7	7	**7**	13	13	**13**	9	9	**9**

Table 8.C.22 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level II

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**2**	–	N/A	**14**	0.38	N/A	**2**	–	N/A	**5**	**–**	N/A
Mathematics	2	**2**	N/A	14	**14**	N/A	2	**2**	N/A	5	**5**	N/A
Science	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**70**	0.77	N/A	**17**	0.63	N/A	**55**	0.79	N/A	**43**	0.55	N/A
Mathematics	70	**70**	N/A	17	**17**	N/A	55	**55**	N/A	43	**43**	N/A
Science	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 8.C.23 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level III

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	0	–	–	**11**	0.32	–	0	–	–	**6**	-	**–**
Mathematics	0	0	–	11	**11**	–	0	0	–	6	**6**	**–**
Science	0	0	0	2	2	**2**	0	0	0	3	3	**3**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**57**	0.69	0.87	**11**	0.73	**–**	**45**	0.83	0.78	**23**	0.63	**–**
Mathematics	57	**57**	0.56	11	**11**	**–**	44	**44**	0.84	23	**23**	**–**
Science	22	22	**22**	6	6	**6**	18	17	**18**	9	9	**9**

Table 8.C.24 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level IV

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**2**	–	–	**18**	0.18	**–**	0	–	–	**7**	**–**	–
Mathematics	2	**2**	–	17	**17**	**–**	0	0	–	7	**7**	–
Science	1	1	**1**	5	5	**5**	0	0	0	0	0	**0**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**67**	0.76	0.57	**19**	0.56	**–**	**49**	0.78	**–**	**53**	0.68	0.93
Mathematics	67	**67**	0.57	19	**19**	**–**	49	**49**	**–**	53	**54**	0.82
Science	19	19	**19**	7	7	**7**	10	10	**10**	14	14	**14**

Table 8.C.25 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level V

	American Indian			Asian			Pacific Islander			Filipino		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**3**	**–**	–	**15**	0.79	**–**	**2**	**–**	–	**12**	0.83	**–**
Mathematics	3	**3**	–	15	**15**	**–**	2	**2**	–	12	**12**	**–**
Science	0	0	**0**	6	6	**6**	1	1	**1**	3	3	**3**
	Hispanic			**African American**			**White**			**Unknown Ethnicity**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**137**	0.65	0.46	**43**	0.64	**–**	**92**	0.63	0.91	**53**	0.63	0.78
Mathematics	137	**137**	0.81	42	**42**	**–**	92	**92**	0.80	53	**53**	0.53
Science	36	36	**36**	6	6	**6**	22	22	**22**	16	16	**16**

Table 8.C.26 CAPA Content Area Correlations by Economic Status: Level I

	Disadvantaged			Not Disadvantaged			Unknown Status		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**8,135**	0.79	0.82	**5,299**	0.79	0.82	**285**	0.73	0.80
Mathematics	8,110	**8,115**	0.77	5,285	**5,289**	0.78	285	**285**	0.72
Science	2,124	2,124	**2,124**	1,323	1,324	**1,326**	62	62	**62**

Table 8.C.27 CAPA Content Area Correlations by Economic Status: Level II

	Disadvantaged			Not Disadvantaged			Unknown Status		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**4,472**	0.74	N/A	**1,963**	0.75	N/A	**208**	0.70	N/A
Mathematics	4,459	**4,459**	N/A	1,955	**1,957**	N/A	208	**208**	N/A
Science	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 8.C.28 CAPA Content Area Correlations by Economic Status: Level III

	Disadvantaged			Not Disadvantaged			Unknown Status		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**4,814**	0.71	0.67	**2,145**	0.75	0.73	**153**	0.68	0.77
Mathematics	4,805	**4,806**	0.67	2,138	**2,140**	0.70	152	**152**	0.69
Science	2,284	2,284	**2,285**	1,045	1,045	**1,046**	60	59	**60**

Table 8.C.29 CAPA Content Area Correlations by Economic Status: Level IV

	Disadvantaged			Not Disadvantaged			Unknown Status		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**6,584**	0.76	0.71	**3,059**	0.77	0.73	**215**	0.71	0.76
Mathematics	6,571	**6,574**	0.71	3,053	**3,056**	0.72	214	**215**	0.69
Science	2,157	2,155	**2,158**	941	940	**941**	56	56	**56**

Table 8.C.30 CAPA Content Area Correlations by Economic Status: Level V

	Disadvantaged			Not Disadvantaged			Unknown Status		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**6,288**	0.69	0.68	**3,572**	0.71	0.72	**357**	0.65	0.70
Mathematics	6,265	**6,272**	0.69	3,559	**3,568**	0.71	356	**356**	0.79
Science	2,059	2,059	**2,063**	1,090	1,091	**1,092**	90	90	**90**

Table 8.C.31 CAPA Content Area Correlations by Disability: Level I

	Mental Retardation			Hard of Hearing			Deafness			Speech Impairment		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**5,197**	0.74	0.79	**85**	0.72	0.58	**47**	0.68	0.56	**116**	0.69	**–**
Mathematics	5,185	**5,188**	0.72	85	**85**	0.65	47	**47**	0.60	116	**116**	**–**
Science	1,392	1,391	**1,392**	25	25	**25**	12	12	**12**	7	7	**7**
	Visual Impairment			**Emotional Disturbance**			**Orthopedic Impairment**			**Other Health Impairment**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**291**	0.83	0.84	**10**	**–**	**–**	**2,356**	0.83	0.83	**330**	0.83	0.87
Mathematics	291	**292**	0.85	10	**10**	**–**	2,348	**2,350**	0.82	328	**329**	0.81
Science	74	74	**74**	1	1	**1**	701	702	**702**	67	67	**67**
	Specific Learning Disability			**Deaf-Blindness**			**Multiple Disabilities**			**Autism**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**106**	0.65	0.59	**28**	0.75	0.97	**1,352**	0.84	0.85	**3,551**	0.71	0.76
Mathematics	104	**104**	0.63	27	**27**	0.84	1,346	**1,348**	0.83	3,546	**3,546**	0.71
Science	16	16	**16**	12	12	**12**	356	357	**357**	804	804	**805**
	Traumatic Brain Injury			**Unknown Disability**								
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**						
ELA	**85**	0.85	0.87	**165**	0.78	0.71						
Mathematics	84	**84**	0.83	163	**163**	0.67						
Science	22	22	**22**	20	20	**20**						

Table 8.C.32 CAPA Content Area Correlations by Disability: Level II

	Mental Retardation			Hard of Hearing			Deafness			Speech Impairment		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**2,091**	0.72	N/A	**35**	0.52	N/A	**40**	0.79	N/A	**650**	0.61	N/A
Mathematics	2,085	**2,086**	N/A	35	**35**	N/A	40	**40**	N/A	648	**648**	N/A
Science	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Visual Impairment			**Emotional Disturbance**			**Orthopedic Impairment**			**Other Health Impairment**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**36**	0.82	N/A	**23**	0.55	N/A	**276**	0.72	N/A	**358**	0.70	N/A
Mathematics	36	**36**	N/A	23	**23**	N/A	275	**275**	N/A	358	**358**	N/A
Science	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Specific Learning Disability			**Deaf-Blindness**			**Multiple Disabilities**			**Autism**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**538**	0.53	N/A	**4**	**–**	N/A	**110**	0.83	N/A	**2,317**	0.78	N/A
Mathematics	538	**538**	N/A	4	**4**	N/A	110	**111**	N/A	2,306	**2,306**	N/A
Science	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Traumatic Brain Injury			**Unknown Disability**								
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**						
ELA	**40**	0.81	N/A	**125**	0.73	N/A						
Mathematics	40	**40**	N/A	124	**124**	N/A						
Science	N/A	N/A	N/A	N/A	N/A	N/A						

Table 8.C.33 CAPA Content Area Correlations by Disability: Level III

	Mental Retardation			Hard of Hearing			Deafness			Speech Impairment		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**2,643**	**0.70**	**0.66**	**61**	**0.76**	**0.76**	**62**	**0.80**	**0.82**	**432**	**0.58**	**0.43**
Mathematics	**2,639**	**2,640**	**0.68**	**61**	**61**	**0.83**	**62**	**62**	**0.86**	**431**	**431**	**0.51**
Science	**1,312**	**1,312**	**1,313**	**28**	**28**	**28**	**29**	**29**	**29**	**162**	**162**	**162**
	Visual Impairment			**Emotional Disturbance**			**Orthopedic Impairment**			**Other Health Impairment**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**42**	0.84	0.83	**39**	0.52	0.82	**344**	0.74	0.55	**351**	0.64	0.57
Mathematics	42	**42**	0.84	39	**39**	0.54	343	**344**	0.57	349	**349**	0.69
Science	28	28	**28**	21	21	**21**	173	174	**174**	168	168	**168**
	Specific Learning Disability			**Deaf-Blindness**			**Multiple Disabilities**			**Autism**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**680**	0.54	0.50	**3**	**–**	**–**	**160**	0.67	0.75	**2,170**	0.76	0.76
Mathematics	678	**679**	0.46	3	**3**	**–**	159	**159**	0.68	2,164	**2,164**	0.72
Science	329	328	**329**	3	3	**3**	86	85	**86**	995	995	**995**
	Traumatic Brain Injury			**Unknown Disability**								
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**						
ELA	**35**	0.73	0.80	**90**	0.58	0.74						
Mathematics	35	**35**	0.49	90	**90**	0.49						
Science	16	16	**16**	39	39	**39**						

Table 8.C.34 CAPA Content Area Correlations by Disability: Level IV

	Mental Retardation			Hard of Hearing			Deafness			Speech Impairment		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**4,434**	0.77	0.72	**53**	0.53	0.67	**93**	0.62	0.50	**310**	0.66	0.65
Mathematics	4,429	**4,430**	0.73	53	**53**	0.85	93	**93**	0.63	310	**310**	0.47
Science	1,511	1,510	**1,512**	24	24	**24**	26	26	**26**	71	71	**71**
	Visual Impairment			**Emotional Disturbance**			**Orthopedic Impairment**			**Other Health Impairment**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**58**	0.69	0.80	**91**	0.61	0.32	**538**	0.77	0.74	**417**	0.73	0.74
Mathematics	58	**58**	0.86	89	**90**	0.48	537	**538**	0.78	417	**418**	0.69
Science	22	22	**22**	26	26	**26**	172	172	**172**	136	136	**136**
	Specific Learning Disability			**Deaf-Blindness**			**Multiple Disabilities**			**Autism**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**732**	0.59	0.48	**4**	**–**	**–**	**280**	0.73	0.60	**2,650**	0.77	0.73
Mathematics	731	**733**	0.41	4	**4**	**–**	278	**278**	0.73	2,644	**2,645**	0.74
Science	223	223	**223**	1	1	**1**	83	83	**83**	802	801	**802**
	Traumatic Brain Injury			**Unknown Disability**								
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**						
ELA	**65**	0.67	0.73	**133**	0.68	0.84						
Mathematics	64	**64**	0.51	131	**131**	0.78						
Science	24	24	**24**	33	32	**33**						

Table 8.C.35 CAPA Content Area Correlations by Disability: Level V

	Mental Retardation			Hard of Hearing			Deafness			Speech Impairment		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**4,976**	0.70	0.69	**64**	0.73	0.80	**115**	0.58	0.60	**207**	0.41	0.28
Mathematics	4,961	**4,966**	0.69	64	**64**	0.80	115	**115**	0.49	206	**206**	0.45
Science	1,602	1,601	**1,605**	22	22	**22**	37	37	**37**	58	58	**58**
	Visual Impairment			**Emotional Disturbance**			**Orthopedic Impairment**			**Other Health Impairment**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**70**	0.71	0.63	**140**	0.64	0.65	**526**	0.73	0.82	**550**	0.66	0.58
Mathematics	70	**70**	0.69	137	**137**	0.46	523	**524**	0.84	546	**548**	0.65
Science	24	24	**24**	54	53	**54**	165	165	**165**	166	166	**166**
	Specific Learning Disability			**Deaf-Blindness**			**Multiple Disabilities**			**Autism**		
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
ELA	**809**	0.46	0.40	**1**	**–**	**–**	**333**	0.76	0.87	**2,155**	0.74	0.71
Mathematics	808	**811**	0.55	1	**1**	**–**	331	**333**	0.78	2,148	**2,151**	0.71
Science	270	271	**271**	0	0	**0**	94	95	**95**	677	678	**678**
	Traumatic Brain Injury			**Unknown Disability**								
	ELA	**Math**	**Science**	**ELA**	**Math**	**Science**						
ELA	**76**	0.62	0.83	**195**	0.66	0.69						
Mathematics	76	**76**	0.61	194	**194**	0.64						
Science	24	24	**24**	46	46	**46**						

Table 8.C.36 Interrater Agreement Analyses for Operational Tasks: Level I

Level I		First Rating			Second Rating			% Agreement			MAD *	Corr †
Content Area	Task	N	Mean	SD	N	Mean	SD	Exact	Adjacent	Neither		
English–Language Arts	1	1,979	4.41	1.39	1,979	4.39	1.40	96.41	2.48	1.11	0.07	0.95
	3	1,979	2.88	1.95	1,979	2.86	1.95	95.76	2.98	1.26	0.07	0.98
	4	1,979	3.75	1.81	1,979	3.72	1.82	95.55	3.13	1.31	0.07	0.97
	6	1,979	2.98	1.92	1,979	2.96	1.93	94.09	4.35	1.57	0.10	0.97
	7	1979	3.71	1.84	1,979	3.70	1.85	96.01	2.93	1.06	0.07	0.98
	9	1,979	3.18	1.94	1,979	3.16	1.94	94.59	3.54	1.87	0.09	0.97
	10	1,979	3.21	1.95	1,979	3.19	1.94	96.21	2.98	0.81	0.06	0.98
	12	1,979	3.10	1.95	1,979	3.08	1.96	94.24	3.44	2.32	0.12	0.95
Mathematics	1	1,949	2.80	1.91	1,949	2.78	1.91	95.43	3.13	1.44	0.08	0.97
	3	1,949	2.92	1.93	1,949	2.91	1.93	93.69	4.21	2.11	0.11	0.96
	4	1,949	2.98	1.93	1,949	2.95	1.92	95.74	2.62	1.64	0.08	0.97
	6	1,949	3.04	2.05	1,949	3.02	2.05	93.74	4.57	1.69	0.11	0.97
	7	1,949	3.04	1.93	1,949	3.04	1.93	94.51	3.64	1.84	0.10	0.97
	9	1,949	3.12	1.93	1,949	3.11	1.93	95.02	3.13	1.84	0.09	0.97
	10	1,949	3.50	1.88	1,949	3.50	1.88	95.43	3.34	1.24	0.08	0.97
	12	1,949	3.18	1.94	1,949	3.20	1.93	93.59	4.26	2.16	0.13	0.94
Science	1	483	3.17	1.95	483	3.13	1.96	95.45	3.11	1.45	0.08	0.97
	3	483	2.86	1.96	483	2.81	1.96	94.41	3.93	1.66	0.10	0.96
	4	483	2.92	1.94	483	2.91	1.95	94.41	3.73	1.87	0.10	0.97
	6	483	3.02	1.95	483	3.02	1.96	94.82	3.31	1.87	0.10	0.96
	7	483	2.92	1.94	483	2.89	1.94	94.20	4.35	1.45	0.09	0.97
	9	483	2.96	1.96	483	3.00	1.97	93.79	4.97	1.25	0.10	0.97
	10	483	3.76	1.83	483	3.76	1.82	94.00	4.55	1.45	0.09	0.97
	12	483	2.77	1.95	483	2.76	1.95	92.55	4.76	2.69	0.15	0.94

* Mean absolute difference between first and second ratings

† Pearson correlation between first and second ratings

Table 8.C.37 Interrater Agreement Analyses for Operational Tasks: Level II

Level II		First Rating			Second Rating			% Agreement			MAD *	Corr †
Content Area	Task	N	Mean	SD	N	Mean	SD	Exact	Adjacent	Neither		
English–Language Arts	1	1,372	2.54	1.39	1,372	2.53	1.39	96.94	1.90	1.17	0.05	0.97
	3	1,372	2.34	1.20	1,372	2.34	1.20	93.15	5.54	1.31	0.09	0.96
	4	1,372	3.60	0.79	1,372	3.60	0.79	96.65	2.55	0.81	0.05	0.93
	6	1,372	2.65	1.22	1,372	2.66	1.22	96.43	2.99	0.59	0.04	0.98
	7	1,372	2.25	1.14	1,372	2.26	1.14	95.55	4.08	0.36	0.05	0.98
	9	1,372	2.30	1.08	1,372	2.30	1.08	94.75	5.03	0.22	0.06	0.97
	10	1,372	1.90	0.92	1,372	1.90	0.92	95.12	4.66	0.22	0.05	0.97
	12	1,372	2.22	1.20	1,372	2.23	1.20	94.97	4.15	0.87	0.06	0.97
Mathematics	1	1,358	3.27	0.83	1,358	3.27	0.84	97.79	1.91	0.29	0.03	0.97
	3	1,358	3.57	0.94	1,358	3.56	0.95	98.60	0.96	0.44	0.02	0.97
	4	1,358	2.04	1.38	1,358	2.03	1.38	96.76	1.99	1.25	0.05	0.98
	6	1,358	1.29	0.82	1,358	1.28	0.82	95.51	3.53	0.95	0.06	0.92
	7	1,358	2.23	1.15	1,358	2.23	1.16	96.91	2.72	0.37	0.04	0.98
	9	1,358	2.39	1.37	1,358	2.38	1.38	96.76	2.58	0.66	0.04	0.98
	10	1,358	2.70	1.29	1,358	2.71	1.28	96.32	2.65	1.03	0.06	0.97
	12	1,358	2.70	1.25	1,358	2.71	1.25	95.66	3.53	0.81	0.06	0.97

* Mean absolute difference between first and second ratings

† Pearson correlation between first and second ratings

Table 8.C.38 Interrater Agreement Analyses for Operational Tasks: Level III

Level III		First Rating			Second Rating			% Agreement			MAD *	Corr †
Content Area	Task	N	Mean	SD	N	Mean	SD	Exact	Adjacent	Neither		
English–Language Arts	1	1,401	3.68	0.73	1,401	3.67	0.73	98.00	1.71	0.28	0.02	0.97
	3	1,401	2.66	1.07	1,401	2.65	1.07	95.07	4.57	0.36	0.05	0.97
	4	1,401	3.02	0.98	1,401	3.02	0.99	95.93	3.35	0.71	0.05	0.95
	6	1,401	2.25	0.90	1,401	2.25	0.89	95.65	3.85	0.50	0.05	0.96
	7	1,401	2.51	1.26	1,401	2.49	1.26	94.79	4.43	0.78	0.06	0.97
	9	1,401	3.19	1.06	1,401	3.18	1.07	96.86	2.57	0.57	0.04	0.96
	10	1,401	2.26	1.19	1,401	2.25	1.19	94.15	5.42	0.43	0.06	0.97
	12	1,401	3.52	0.90	1,401	3.56	0.83	96.79	1.71	1.50	0.07	0.86
Mathematics	1	1,399	2.12	1.18	1,399	2.13	1.18	96.71	2.72	0.58	0.04	0.98
	3	1,399	1.97	1.09	1,399	1.97	1.08	92.78	6.15	1.07	0.09	0.94
	4	1,399	3.01	1.27	1,399	2.99	1.28	97.28	2.29	0.42	0.04	0.98
	6	1,399	2.23	0.92	1,399	2.23	0.93	92.85	6.43	0.71	0.08	0.94
	7	1,399	2.39	1.03	1,399	2.40	1.04	96.14	3.50	0.35	0.04	0.97
	9	1,399	2.79	0.83	1,399	2.79	0.84	96.64	2.72	0.64	0.04	0.96
	10	1,399	1.66	1.02	1,399	1.66	1.02	97.71	1.79	0.50	0.03	0.97
	12	1,399	2.29	1.14	1,399	2.29	1.14	94.92	4.00	1.07	0.07	0.96
Science	1	660	2.52	1.04	660	2.53	1.06	94.24	4.09	1.66	0.08	0.94
	3	660	2.98	0.89	660	3.00	0.87	95.15	4.39	0.45	0.06	0.95
	4	660	2.67	1.15	660	2.65	1.16	95.15	4.39	0.45	0.05	0.98
	6	660	2.48	1.38	660	2.47	1.39	97.12	1.97	0.91	0.05	0.97
	7	660	2.74	1.04	660	2.72	1.05	96.06	3.18	0.75	0.05	0.96
	9	660	2.76	1.17	660	2.74	1.20	96.36	2.73	0.90	0.06	0.96
	10	660	2.30	1.11	660	2.30	1.13	95.91	2.73	1.36	0.06	0.96
	12	660	2.97	1.06	660	2.96	1.08	94.85	4.24	0.90	0.07	0.95

* Mean absolute difference between first and second ratings

† Pearson correlation between first and second ratings

Table 8.C.39 Interrater Agreement Analyses for Operational Tasks: Level IV

Level IV		First Rating			Second Rating			% Agreement			MAD *	Corr †
Content Area	Task	N	Mean	SD	N	Mean	SD	Exact	Adjacent	Neither		
English–Language Arts	1	1,481	2.82	1.33	1,481	2.83	1.33	96.89	2.90	0.21	0.03	0.99
	3	1,481	2.45	1.16	1,481	2.45	1.16	92.64	6.68	0.68	0.08	0.96
	4	1,481	1.50	1.05	1,481	1.51	1.04	93.25	5.60	1.15	0.09	0.94
	6	1,481	2.61	1.17	1,481	2.61	1.16	91.49	7.83	0.68	0.10	0.95
	7	1,481	3.54	0.78	1,481	3.54	0.79	97.50	2.09	0.41	0.03	0.96
	9	1,481	2.62	1.14	1,481	2.62	1.14	92.17	6.89	0.95	0.09	0.95
	10	1,481	2.57	1.07	1,481	2.57	1.07	95.81	3.38	0.82	0.06	0.96
	12	1,481	2.15	1.24	1,481	2.17	1.23	95.07	4.12	0.82	0.06	0.97
Mathematics	1	1,478	2.34	1.38	1,478	2.35	1.38	97.56	1.89	0.55	0.03	0.98
	3	1,478	2.94	1.29	1,478	2.94	1.30	96.68	2.57	0.74	0.04	0.98
	4	1,478	1.76	1.27	1478	1.77	1.27	97.23	1.83	0.95	0.04	0.97
	6	1,478	2.04	1.22	1,478	2.04	1.22	94.32	3.79	1.90	0.09	0.94
	7	1,478	2.99	1.05	1,478	2.99	1.05	96.48	3.25	0.27	0.04	0.98
	9	1,478	2.95	1.29	1478	2.95	1.29	97.16	2.10	0.74	0.04	0.97
	10	1,478	1.83	0.98	1,478	1.83	0.99	95.20	4.19	0.61	0.06	0.96
	12	1,478	2.64	1.06	1,478	2.64	1.06	91.27	7.37	1.35	0.10	0.93
Science	1	399	3.01	1.09	399	3.01	1.09	97.49	2.51	0.00	0.03	0.99
	3	399	2.47	1.02	399	2.47	1.02	94.24	5.51	0.25	0.06	0.97
	4	399	2.32	1.08	399	2.33	1.08	92.98	5.76	1.25	0.09	0.95
	6	399	3.10	1.00	399	3.11	1.00	93.73	5.01	1.25	0.08	0.93
	7	399	3.04	0.79	399	3.05	0.78	93.98	5.76	0.25	0.06	0.95
	9	399	2.62	0.95	399	2.62	0.97	95.24	4.26	0.50	0.05	0.97
	10	399	2.35	1.02	399	2.36	1.02	91.98	7.27	0.75	0.09	0.94
	12	399	2.98	1.03	399	2.98	1.04	94.99	4.26	0.75	0.06	0.97

* Mean absolute difference between first and second ratings

† Pearson correlation between first and second ratings

Table 8.C.40 Interrater Agreement Analyses for Operational Tasks: Level V

Level V		First Rating			Second Rating			% Agreement			MAD *	Corr †
Content Area	Task	N	Mean	SD	N	Mean	SD	Exact	Adjacent	Neither		
English–Language Arts	1	1,083	3.15	0.80	1,083	3.15	0.81	96.12	3.42	0.46	0.04	0.96
	3	1,083	2.60	1.05	1,083	2.60	1.06	91.23	8.03	0.74	0.10	0.95
	4	1,083	2.78	1.05	1,083	2.77	1.05	93.26	5.45	1.30	0.09	0.93
	6	1,083	1.90	1.10	1,083	1.89	1.10	91.87	6.28	1.84	0.10	0.94
	7	1,083	3.23	1.08	1,083	3.24	1.08	93.63	5.82	0.55	0.07	0.96
	9	1,083	2.48	1.16	1,083	2.47	1.17	91.69	7.66	0.64	0.09	0.96
	10	1,083	2.48	1.21	1,083	2.47	1.22	92.71	6.65	0.65	0.08	0.97
	12	1,083	3.55	0.86	1,083	3.57	0.82	95.29	3.51	1.20	0.08	0.87
Mathematics	1	1,076	3.41	1.14	1,076	3.41	1.14	98.14	1.58	0.28	0.03	0.98
	3	1,076	2.79	1.09	1,076	2.80	1.10	95.26	3.72	1.02	0.06	0.96
	4	1,076	2.32	1.36	1,076	2.32	1.36	96.10	2.79	1.11	0.06	0.97
	6	1,076	2.35	1.21	1,076	2.34	1.22	95.82	3.07	1.11	0.05	0.97
	7	1,076	2.42	1.21	1,076	2.42	1.21	93.87	5.20	0.92	0.07	0.96
	9	1,076	2.55	1.42	1,076	2.55	1.41	96.38	2.79	0.84	0.05	0.97
	10	1,076	2.01	1.17	1,076	2.02	1.18	93.12	5.67	1.21	0.08	0.96
	12	1,076	2.18	1.24	1,076	2.19	1.23	93.40	4.83	1.77	0.09	0.95
Science	1	353	2.40	0.88	353	2.42	0.91	91.50	8.50	0.00	0.08	0.95
	3	353	2.36	0.96	353	2.36	0.97	95.18	4.82	0.00	0.05	0.97
	4	353	2.75	0.94	353	2.73	0.95	95.18	4.53	0.28	0.06	0.95
	6	353	2.18	1.10	353	2.17	1.13	91.78	7.65	0.57	0.09	0.96
	7	353	2.56	1.04	353	2.56	1.03	90.65	7.37	1.98	0.12	0.92
	9	353	2.80	0.82	353	2.79	0.81	94.05	4.82	1.14	0.08	0.91
	10	353	3.40	0.92	353	3.40	0.94	95.47	3.40	1.13	0.06	0.93
	12	353	2.75	1.07	353	2.76	1.07	90.65	8.22	1.13	0.12	0.91

* Mean absolute difference between first and second ratings

† Pearson correlation between first and second ratings

Appendix 8.D—IRT Analyses

Table 8.D.1 Item Classifications for Model-Data Fit Across All CAPA Levels

Fit Classification	ELA No. of Items	Mathematics No. of Items	Science No. of Items
A	70	89	53
B	36	26	8
C	13	5	3
D	1	0	0
F	0	0	0

Table 8.D.2 Fit Classifications: Level I Tasks

Fit	ELA Frequency	Mathematics Frequency	Science Frequency
A	8	19	12
B	12	5	3
C	4	0	1
D	0	0	0
F	0	0	0

Table 8.D.3 Fit Classifications: Level II Tasks

Fit	ELA Frequency	Mathematics Frequency
A	16	19
B	8	4
C	0	1
D	0	0
F	0	0

Table 8.D.4 Fit Classifications: Level III Tasks

Fit	ELA Frequency	Mathematics Frequency	Science Frequency
A	17	21	12
B	5	3	3
C	2	0	1
D	0	0	0
F	0	0	0

Table 8.D.5 Fit Classifications: Level IV Tasks

Fit	ELA Frequency	Mathematics Frequency	Science Frequency
A	19	13	14
B	5	8	1
C	0	3	1
D	0	0	0
F	0	0	0

Table 8.D.6 Fit Classifications: Level V Tasks

Fit	ELA Frequency	Mathematics Frequency	Science Frequency
A	10	17	15
B	6	6	1
C	7	1	0
D	1	0	0
F	0	0	0

Table 8.D.7 IRT *b*-values for ELA, by Level

Level		Number of Items	Mean	Standard Deviation	Min	Max
I	All Operational Items	8	–0.61	0.27	–1.18	–0.40
	Field-Test Items	16	–0.68	0.18	–1.06	–0.42
II	All Operational Items	8	–0.99	0.60	–2.18	–0.28
	Field-Test Items	16	–1.19	0.75	–2.39	0.13
III	All Operational Items	8	–1.32	0.89	–2.77	–0.30
	Field-Test Items	16	–1.39	0.67	–2.49	–0.57
IV	All Operational Items	8	–1.05	0.64	–2.30	0.01
	Field-Test Items	16	–1.10	0.47	–1.88	0.00
V	All Operational Items	8	–1.16	0.68	–2.19	–0.22
	Field-Test Items	16	–1.32	0.41	–1.89	–0.52

Table 8.D.8 IRT *b*-values for Mathematics, by Level

Level		Number of Items	Mean	Standard Deviation	Min	Max
I	All Operational Items	8	–0.21	0.12	–0.43	0.00
	Field-test Items	16	–0.20	0.17	–0.40	0.11
II	All Operational Items	8	–1.02	0.68	–1.91	0.18
	Field-test Items	16	–1.24	0.71	–2.19	0.53
III	All Operational Items	8	–0.87	0.30	–1.33	–0.50
	Field-test Items	16	–1.18	0.44	–2.13	–0.56
IV	All Operational Items	8	–0.86	0.48	–1.47	–0.05
	Field-test Items	16	–1.09	0.67	–2.06	0.50
V	All Operational Items	8	–0.98	0.34	–1.54	–0.43
	Field-test Items	16	–1.36	0.31	–1.78	–0.88

Table 8.D.9 IRT *b*-values for Science, by Level

Level		Number of Items	Mean	Standard Deviation	Min	Max
I	All Operational Items	8	–0.33	0.14	–0.64	–0.18
	Field-test Items	8	–0.48	0.25	–0.84	–0.08
III	All Operational Items	8	–1.29	0.33	–1.90	–0.88
	Field-test Items	8	–1.04	0.25	–1.35	–0.58
IV	All Operational Items	8	–1.19	0.30	–1.56	–0.80
	Field-test Items	8	–0.99	0.52	–1.96	–0.38
V	All Operational Items	8	–0.52	0.40	–1.32	–0.09
	Field-test Items	8	–0.07	0.33	–0.47	0.47

Table 8.D.10 Score Conversions: Level I, ELA

Raw Score	Theta	Scale Score	CSEM	Performance Level
40	N/A	60	0	**Advanced**
39	1.1923	54	6	
38	0.8284	50	6	
37	0.6418	48	5	
36	0.5152	47	4	
35	0.4180	46	3	
34	0.3379	45	3	
33	0.2688	44	3	
32	0.2072	43	3	
31	0.1511	43	3	
30	0.0989	42	3	
29	0.0497	41	3	
28	0.0026	41	2	
27	–0.0429	40	2	
26	–0.0874	40	2	
25	–0.1311	39	2	**Proficient**
24	–0.1746	39	2	
23	–0.2182	38	2	
22	–0.2621	38	2	
21	–0.3067	37	2	
20	–0.3524	37	2	
19	–0.3994	36	3	
18	–0.4483	36	3	
17	–0.4995	35	3	
16	–0.5534	35	3	
15	–0.6108	34	3	**Basic**
14	–0.6723	33	3	
13	–0.7388	32	3	
12	–0.8115	32	3	
11	–0.8919	31	3	
10	–0.9822	30	4	
9	–1.0862	28	4	**Below Basic**
8	–1.2104	27	4	
7	–1.3663	25	5	
6	–1.5735	23	6	
5	–1.8563	20	6	
4	–2.2228	15	4	**Far Below Basic**
3	–2.6675	15	0	
2	–3.2224	15	0	
1	–4.0512	15	0	
0	N/A	15	0	

Table 8.D.11 Score Conversions: Level II, ELA

Raw Score	Theta	Scale Score	CSEM	Performance Level
32	N/A	60	0	**Advanced**
31	2.6480	52	6	
30	1.9602	48	4	
29	1.5485	46	3	
28	1.2497	45	3	
27	1.0131	44	2	
26	0.8157	43	2	
25	0.6447	42	2	
24	0.4922	41	2	
23	0.3526	40	2	
22	0.2220	40	2	
21	0.0971	39	2	**Proficient**
20	–0.0247	39	2	
19	–0.1460	38	2	
18	–0.2692	37	2	
17	–0.3970	37	2	
16	–0.5323	36	2	
15	–0.6786	35	2	
14	–0.8393	35	2	
13	–1.0181	34	2	**Basic**
12	–1.2184	33	2	
11	–1.4430	32	2	
10	–1.6952	30	3	
9	–1.9801	29	3	**Below Basic**
8	–2.3039	27	3	
7	–2.6678	26	3	
6	–3.0639	24	3	
5	–3.4823	22	3	
4	–3.9241	20	3	
3	–4.4096	17	3	**Far Below Basic**
2	–4.9931	15	1	
1	–5.8466	15	0	
0	N/A	15	0	

Table 8.D.12 Score Conversions: Level III, ELA

Raw Score	Theta	Scale Score	CSEM	Performance Level
32	N/A	60	0	**Advanced**
31	2.8958	50	6	
30	2.0892	47	3	
29	1.5767	45	3	
28	1.1914	43	2	
27	0.8797	42	2	
26	0.6154	41	2	
25	0.3828	40	2	
24	0.1714	39	2	**Proficient**
23	–0.0258	39	2	
22	–0.2136	38	2	
21	–0.3952	37	2	
20	–0.5724	37	2	
19	–0.7463	36	2	
18	–0.9177	35	2	
17	–1.0867	35	2	
16	–1.2540	34	2	**Basic**
15	–1.4201	33	2	
14	–1.5862	33	2	
13	–1.7539	32	2	
12	–1.9253	31	2	
11	–2.1034	31	2	
10	–2.2917	30	2	
9	–2.4947	29	2	**Below Basic**
8	–2.7179	28	2	
7	–2.9675	27	2	
6	–3.2511	26	2	
5	–3.5771	25	2	
4	–3.9575	24	2	
3	–4.4140	22	3	**Far Below Basic**
2	–4.9976	20	3	
1	–5.8759	16	2	
0	N/A	15	0	

Table 8.D.13 Score Conversions: Level IV, ELA

Raw Score	Theta	Scale Score	CSEM	Performance Level
32	N/A	60	0	**Advanced**
31	2.4856	54	5	
30	1.8518	50	4	
29	1.4845	48	3	
28	1.2158	46	3	
27	0.9964	45	3	
26	0.8061	44	2	
25	0.6348	43	2	
24	0.4771	42	2	
23	0.3290	41	2	**Proficient**
22	0.1881	40	2	
21	0.0519	40	2	
20	–0.0817	39	2	
19	–0.2147	38	2	
18	–0.3492	37	2	
17	–0.4878	36	2	
16	–0.6331	36	2	
15	–0.7884	35	2	
14	–0.9573	34	2	**Basic**
13	–1.1430	33	3	
12	–1.3482	31	3	
11	–1.5739	30	3	
10	–1.8204	29	3	**Below Basic**
9	–2.0890	27	3	
8	–2.3848	25	3	
7	–2.7136	23	3	
6	–3.0782	21	4	
5	–3.4758	19	4	
4	–3.9067	16	3	**Far Below Basic**
3	–4.3867	15	1	
2	–4.9668	15	0	
1	–5.8173	15	0	
0	N/A	15	0	

Table 8.D.14 Score Conversions: Level V, ELA

Raw Score	Theta	Scale Score	CSEM	Performance Level
32	N/A	60	0	**Advanced**
31	2.9856	49	6	
30	2.2045	46	3	
29	1.7235	45	2	
28	1.3677	43	2	
27	1.0790	42	2	
26	0.8306	41	2	
25	0.6077	40	2	
24	0.4020	40	2	
23	0.2082	39	2	**Proficient**
22	0.0228	38	2	
21	–0.1567	38	2	
20	–0.3323	37	2	
19	–0.5056	36	2	
18	–0.6779	36	2	
17	–0.8507	35	2	
16	–1.0251	34	2	**Basic**
15	–1.2027	34	2	
14	–1.3847	33	2	
13	–1.5726	32	2	
12	–1.7676	31	2	
11	–1.9716	31	2	
10	–2.1864	30	2	
9	–2.4149	29	2	**Below Basic**
8	–2.6605	28	2	
7	–2.9269	27	2	
6	–3.2184	26	2	
5	–3.5406	25	2	
4	–3.9036	23	2	
3	–4.3283	22	3	**Far Below Basic**
2	–4.8660	20	3	
1	–5.6853	17	2	
0	N/A	15	0	

Table 8.D.15 Score Conversions: Level I, Mathematics

Raw Score	Theta	Scale Score	CSEM	Performance Level
40	N/A	60	0	**Advanced**
39	1.5267	49	9	
38	1.1715	46	6	
37	0.9916	44	4	
36	0.8707	43	3	
35	0.7788	42	3	
34	0.7037	41	3	
33	0.6396	41	2	
32	0.5830	40	2	
31	0.5319	40	2	
30	0.4849	39	2	
29	0.4409	39	2	
28	0.3993	38	2	**Proficient**
27	0.3594	38	2	
26	0.3209	38	2	
25	0.2832	37	2	
24	0.2462	37	2	
23	0.2095	36	2	
22	0.1728	36	2	
21	0.1358	36	2	
20	0.0982	35	2	
19	0.0597	35	2	
18	0.0200	35	2	
17	–0.0215	34	2	**Basic**
16	–0.0651	34	2	
15	–0.1117	33	2	
14	–0.1621	33	2	
13	–0.2175	32	2	
12	–0.2797	32	3	
11	–0.3513	31	3	
10	–0.4365	30	3	
9	–0.5426	29	4	**Below Basic**
8	–0.6825	28	4	
7	–0.8802	26	5	
6	–1.1702	23	6	
5	–1.5610	19	6	
4	–2.0134	15	3	**Far Below Basic**
3	–2.5130	15	0	
2	–3.1051	15	0	
1	–3.9620	15	0	
0	N/A	15	0	

Table 8.D.16 Score Conversions: Level II, Mathematics

Raw Score	Theta	Scale Score	CSEM	Performance Level
32	N/A	60	0	**Advanced**
31	2.2361	54	6	
30	1.7119	50	5	
29	1.3878	47	4	
28	1.1403	46	3	
27	0.9363	44	3	
26	0.7613	43	3	
25	0.6066	42	3	
24	0.4657	41	3	
23	0.3338	40	3	**Proficient**
22	0.2074	39	3	
21	0.0833	38	3	
20	–0.0408	37	3	
19	–0.1671	36	3	
18	–0.2974	35	3	
17	–0.4332	34	3	**Basic**
16	–0.5757	33	3	
15	–0.7259	32	3	
14	–0.8847	31	3	
13	–1.0535	30	3	
12	–1.2356	28	3	**Below Basic**
11	–1.4367	27	3	
10	–1.6655	25	4	
9	–1.9329	23	4	
8	–2.2471	21	4	
7	–2.6067	18	4	
6	–2.9986	15	3	**Far Below Basic**
5	–3.4104	15	0	
4	–3.8437	15	0	
3	–4.3197	15	0	
2	–4.8933	15	0	
1	–5.7370	15	0	
0	N/A	15	0	

Table 8.D.17 Score Conversions: Level III, Mathematics

Raw Score	Theta	Scale Score	CSEM	Performance Level
32	N/A	60	0	Advanced
31	2.7614	50	7	
30	2.0779	47	4	
29	1.6850	45	3	
28	1.4074	44	3	
27	1.1895	42	2	
26	1.0065	42	2	
25	0.8454	41	2	
24	0.6984	40	2	
23	0.5607	39	2	Proficient
22	0.4290	39	2	
21	0.3010	38	2	
20	0.1748	37	2	
19	0.0487	37	2	
18	–0.0786	36	2	
17	–0.2087	36	2	
16	–0.3435	35	2	
15	–0.4855	34	2	Basic
14	–0.6381	33	2	
13	–0.8063	33	2	
12	–0.9970	32	2	
11	–1.2202	30	2	
10	–1.4890	29	3	Below Basic
9	–1.8182	28	3	
8	–2.2148	26	3	
7	–2.6614	23	3	
6	–3.1215	21	3	
5	–3.5748	19	3	
4	–4.0296	17	3	
3	–4.5153	15	1	Far Below Basic
2	–5.0921	15	0	
1	–5.9356	15	0	
0	N/A	15	0	

Table 8.D.18 Score Conversions: Level IV, Mathematics

Raw Score	Theta	Scale Score	CSEM	Performance Level
32	N/A	60	0	
31	2.5295	54	6	
30	1.8896	49	5	
29	1.5355	47	4	
28	1.2890	45	3	**Advanced**
27	1.0953	44	3	
26	0.9312	43	3	
25	0.7849	42	3	
24	0.6499	41	2	
23	0.5222	40	2	
22	0.3994	39	2	
21	0.2798	38	2	
20	0.1623	37	2	**Proficient**
19	0.0457	37	2	
18	–0.0712	36	2	
17	–0.1895	35	2	
16	–0.3111	34	2	
15	–0.4379	33	2	
14	–0.5733	32	3	**Basic**
13	–0.7216	31	3	
12	–0.8897	30	3	
11	–1.0887	29	3	
10	–1.3359	27	4	
9	–1.6561	25	4	**Below Basic**
8	–2.0690	22	5	
7	–2.5517	19	5	
6	–3.0444	15	3	
5	–3.5187	15	0	
4	–3.9866	15	0	
3	–4.4815	15	0	**Far Below Basic**
2	–5.0655	15	0	
1	–5.9155	15	0	
0	N/A	15	0	

Table 8.D.19 Score Conversions: Level V, Mathematics

Raw Score	Theta	Scale Score	CSEM	Performance Level
32	N/A	60	0	**Advanced**
31	2.0904	49	8	
30	1.5551	46	4	
29	1.2540	44	3	
28	1.0401	43	3	
27	0.8705	42	2	
26	0.7271	41	2	
25	0.6003	40	2	
24	0.4846	40	2	
23	0.3761	39	2	**Proficient**
22	0.2723	38	2	
21	0.1709	38	2	
20	0.0703	37	2	
19	–0.0314	36	2	
18	–0.1356	36	2	
17	–0.2441	35	2	
16	–0.3587	34	2	**Basic**
15	–0.4811	34	2	
14	–0.6136	33	2	
13	–0.7593	32	2	
12	–0.9234	31	3	
11	–1.1159	30	3	
10	–1.3562	28	3	**Below Basic**
9	–1.6797	26	4	
8	–2.1180	24	4	
7	–2.6249	21	4	
6	–3.1199	18	4	
5	–3.5869	15	2	**Far Below Basic**
4	–4.0466	15	0	
3	–4.5341	15	0	
2	–5.1117	15	0	
1	–5.9556	15	0	
0	N/A	15	0	

Table 8.D.20 Score Conversions: Level I, Science

Raw Score	Theta	Scale Score	CSEM	Performance Level
40	N/A	60	0	
39	1.5037	49	9	
38	1.1424	46	6	
37	0.9576	44	4	
36	0.8328	43	3	
35	0.7375	42	3	**Advanced**
34	0.6594	41	3	
33	0.5924	40	3	
32	0.5332	40	2	
31	0.4794	39	2	
30	0.4298	39	2	
29	0.3833	38	2	
28	0.3391	38	2	
27	0.2966	37	2	
26	0.2554	37	2	
25	0.2150	36	2	**Proficient**
24	0.1752	36	2	
23	0.1356	36	2	
22	0.0959	35	2	
21	0.0557	35	2	
20	0.0148	34	2	
19	–0.0272	34	2	
18	–0.0708	34	2	
17	–0.1164	33	2	
16	–0.1646	33	2	**Basic**
15	–0.2162	32	2	
14	–0.2724	32	2	
13	–0.3345	31	3	
12	–0.4048	30	3	
11	–0.4866	29	3	
10	–0.5855	28	3	
9	–0.7111	27	4	
8	–0.8814	25	5	**Below Basic**
7	–1.1258	23	5	
6	–1.4678	20	6	
5	–1.8802	16	4	
4	–2.3243	15	0	
3	–2.8064	15	0	
2	–3.3807	15	0	**Far Below Basic**
1	–4.2214	15	0	
0	N/A	15	0	

Table 8.D.21 Score Conversions: Level III, Science

Raw Score	Theta	Scale Score	CSEM	Performance Level
32	N/A	60	0	
31	2.3419	44	9	
30	1.6730	42	3	**Advanced**
29	1.2825	41	2	
28	1.0023	40	2	
27	0.7806	39	2	
26	0.5942	38	1	
25	0.4306	38	1	
24	0.2822	37	1	
23	0.1441	37	1	**Proficient**
22	0.0123	36	1	
21	–0.1159	36	1	
20	–0.2430	35	1	
19	–0.3710	35	1	
18	–0.5021	34	1	
17	–0.6384	34	1	
16	–0.7825	33	1	
15	–0.9372	33	1	**Basic**
14	–1.1063	32	1	
13	–1.2942	32	2	
12	–1.5065	31	2	
11	–1.7495	30	2	
10	–2.0294	29	2	
9	–2.3497	28	2	
8	–2.7076	27	2	**Below Basic**
7	–3.0936	25	2	
6	–3.4965	24	2	
5	–3.9127	23	2	
4	–4.3506	21	2	
3	–4.8343	19	3	
2	–5.4197	17	3	**Far Below Basic**
1	–6.2790	15	1	
0	N/A	15	0	

Table 8.D.22 Score Conversions: Level IV, Science

Raw Score	Theta	Scale Score	CSEM	Performance Level
32	N/A	60	0	**Advanced**
31	2.9330	47	8	
30	2.1786	44	3	
29	1.7163	42	3	
28	1.3741	41	2	
27	1.0973	40	2	
26	0.8615	39	2	**Proficient**
25	0.6532	38	2	
24	0.4645	37	2	
23	0.2899	37	2	
22	0.1257	36	2	
21	–0.0312	35	2	
20	–0.1834	35	2	
19	–0.3333	34	2	**Basic**
18	–0.4830	34	2	
17	–0.6351	33	2	
16	–0.7922	32	2	
15	–0.9574	32	2	
14	–1.1344	31	2	
13	–1.3281	30	2	
12	–1.5444	29	2	**Below Basic**
11	–1.7901	28	2	
10	–2.0719	27	2	
9	–2.3933	26	2	
8	–2.7507	25	2	
7	–3.1339	23	2	
6	–3.5325	22	2	
5	–3.9442	20	3	
4	–4.3778	18	3	**Far Below Basic**
3	–4.8578	17	2	
2	–5.4402	15	1	
1	–6.2981	15	0	
0	N/A	15	0	

Table 8.D.23 Score Conversions: Level V, Science

Raw Score	Theta	Scale Score	CSEM	Performance Level
32	N/A	60	0	
31	3.6948	47	8	
30	2.9225	44	3	
29	2.4417	42	2	**Advanced**
28	2.0807	41	2	
27	1.7851	40	2	
26	1.5296	39	2	
25	1.3006	38	2	
24	1.0899	38	2	
23	0.8921	37	1	
22	0.7040	36	1	**Proficient**
21	0.5231	36	1	
20	0.3478	35	1	
19	0.1765	35	1	
18	0.0080	34	1	
17	–0.1591	33	1	
16	–0.3265	33	1	
15	–0.4963	32	1	**Basic**
14	–0.6708	32	1	
13	–0.8533	31	1	
12	–1.0474	30	2	
11	–1.2577	30	2	
10	–1.4895	29	2	
9	–1.7475	28	2	
8	–2.0345	27	2	
7	–2.3500	26	2	
6	–2.6905	25	2	**Below Basic**
5	–3.0545	24	2	
4	–3.4484	22	2	
3	–3.8931	21	2	
2	–4.4415	19	3	
1	–5.2644	16	2	**Far Below Basic**
0	N/A	15	0	

Appendix 8.E—DIF Analyses

Table 8.E.1 Item Exhibiting Significant DIF by Ethnic Group

Content Area	Task No.	Level	Task#	Version	SMD	Comparison	In Favor Of
English–Language Arts Operational Tasks	VE093618	IV	1	Operational	0.398	White/Asian	Asian
	VE093618	IV	1	Operational	0.337	White/CombAsian	CombAsian
	VC208668	V	6	Operational	0.321	White/Asian	Asian
	VC208668	V	6	Operational	0.358	White/Filipino	Filipino
	VC208668	V	6	Operational	0.303	White/CombAsian	CombAsian
English–Language Arts Field-test Tasks	VE390331	II	11	1	–0.249	White/CombAsian	White
	VE392482	III	5	1	–0.328	White/CombAsian	White
	VE404044	IV	2	3	0.364	White/Asian	Asian
	VE404044	IV	2	3	0.329	White/CombAsian	CombAsian
	VE436315	V	11	1	–0.283	White/Asian	White
	VE436302	V	11	3	0.398	White/Asian	Asian
	VE436302	V	11	3	0.356	White/CombAsian	CombAsian
Mathematics Operational Tasks	VE098638	III	19	Operational	0.296	White/Asian	Asian
Mathematics Field-test Tasks*	–	–	–	–	–	–	–
Science Operational Tasks*	–	–	–	–	–	–	–
Science Field-test Tasks*	–	–	–	–	–	–	–

* No items exhibited significant ethnic DIF.

Table 8.E.2 Items Exhibiting Significant DIF by Disability Group

Content Area	Task No.	Level	Task#	Version	SMD	Comparison	In Favor Of
English–Language Arts Operational Tasks	VE099155	III	1	Operational	–0.191	MentRetard/MultiDisab	MentRetard
	VE208294	III	9	Operational	0.307	MentRetard/Autism	Autism
	VE093618	IV	1	Operational	0.670	MentRetard/Autism	Autism
	VE476678	IV	3	Operational	–0.318	MentRetard/Autism	MentRetard
	VC208470	IV	6	Operational	–0.404	MentRetard/Autism	MentRetard
	VE093557	IV	12	Operational	0.418	MentRetard/Autism	Autism
	VC208668	V	6	Operational	0.617	MentRetard/Autism	Autism
English–Language Arts Field–test Tasks	VE388128	1	5	1	0.518	MentRetard/OrthoImped	OrthoImped
	VE089782	1	5	2	–0.449	MentRetard/OrthoImped	MentRetard
	VE089892	1	11	3	–0.616	MentRetard/OrthoImped	MentRetard
	VE089892	1	11	3	–0.543	MentRetard/OrthoImped	MentRetard
	VE404044	4	2	3	–0.427	MentRetard/OrthoImped	MentRetard
	VE404080	4	5	1	–0.504	MentRetard/Autism	MentRetard
	VE404065	4	8	1	–0.425	MentRetard/Autism	MentRetard
	VE404040	4	2	2	0.665	MentRetard/Autism	Autism
	VE404068	4	5	2	0.351	MentRetard/Autism	Autism
	VE404056	4	8	2	–0.518	MentRetard/Autism	MentRetard
	VE404044	4	2	3	0.553	MentRetard/Autism	Autism
	VE404070	4	5	3	0.424	MentRetard/Autism	Autism
	VE404081	4	11	4	–0.400	MentRetard/Autism	MentRetard
	VE436302	5	11	3	0.328	MentRetard/SpfcLearn	SpfcLearn
	VE436315	5	11	1	–0.300	MentRetard/Autism	MentRetard
	VE436288	5	11	2	0.498	MentRetard/Autism	Autism
	VE436302	5	11	3	0.913	MentRetard/Autism	Autism
	VE436307	5	5	4	0.583	MentRetard/Autism	Autism
Mathematics Operational Tasks	VC471057	2	18	Operational	–0.266	MentRetard/SpfcLearn	MentRetard
	VC471057	2	18	Operational	–0.218	MentRetard/Autism	MentRetard
	VC207351	3	13	Operational	0.388	MentRetard/SpeechImp	SpeechImp
	VC468572	3	15	Operational	–0.275	MentRetard/SpeechImp	MentRetard
	VC207351	3	13	Operational	0.417	MentRetard/SpfcLearn	SpfcLearn
	VC335667	3	21	Operational	0.212	MentRetard/SpfcLearn	SpfcLearn
	VE098690	3	24	Operational	–0.335	MentRetard/SpfcLearn	MentRetard
	VC208066	5	21	Operational	–0.415	MentRetard/EmoDisturb	MentRetard

Content Area	Task No.	Level	Task#	Version	SMD	Comparison	In Favor Of
Mathematics Field–test Tasks	VC208066	5	21	Operational	–0.454	MentRetard/SpfcLearn	MentRetard
	VE090703	1	14	1	–0.500	MentRetard/OrthoImped	MentRetard
	VC464403	1	20	3	–0.528	MentRetard/OrthoImped	MentRetard
	VC335367	1	20	2	0.642	MentRetard/Autism	Autism
	VE436484	2	23	4	–0.312	MentRetard/Autism	MentRetard
	VC098576	3	14	1	0.384	MentRetard/SpfcLearn	SpfcLearn
	VC335640	3	20	1	0.438	MentRetard/SpfcLearn	SpfcLearn
	VC335620	3	17	2	0.414	MentRetard/Autism	Autism
	VC093066	3	20	2	0.422	MentRetard/Autism	Autism
	VC465936	3	23	3	0.550	MentRetard/Autism	Autism
	VC098584	3	20	4	0.343	MentRetard/Autism	Autism
	VE389965	4	17	1	0.281	MentRetard/SpfcLearn	SpfcLearn
	VE389910	4	20	1	0.255	MentRetard/SpfcLearn	SpfcLearn
	VE099828	5	17	1	0.370	MentRetard/SpfcLearn	SpfcLearn
	VE098885	5	14	3	0.522	MentRetard/SpfcLearn	SpfcLearn
	VE436590	5	17	3	0.698	MentRetard/SpfcLearn	SpfcLearn
	VE099828	5	17	1	0.343	MentRetard/Autism	Autism
	VE436598	5	20	1	0.397	MentRetard/Autism	Autism
	VE098885	5	14	3	0.420	MentRetard/Autism	Autism
	VE436590	5	17	3	0.429	MentRetard/Autism	Autism
Science Operational Tasks	VC206649	4	28	Operational	–0.278	MentRetard/SpfcLearn	MentRetard
Science Field–test Tasks	VE400284	1	29	1,3	–1.167	MentRetard/OrthoImped	MentRetard
	VE400284	1	29	1,3	–0.626	MentRetard/MultiDisab	MentRetard

Table 8.E.3 CAPA Disability Distributions: Level I

Disability	ELA		Mathematics		Science	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Mental retardation	5,197	37.9	5,188	37.9	1,392	39.7
Hard of hearing	85	0.6	85	0.6	25	0.7
Deafness	47	0.3	47	0.3	12	0.3
Speech or language impairment*	116	0.8	116	0.8	–	–
Visual impairment	291	2.1	292	2.1	74	2.1
Emotional disturbance*	–	–	–	–	–	–
Orthopedic impairment	2,356	17.2	2,350	17.2	702	20.0
Other health impairment	330	2.4	329	2.4	67	1.9
Specific learning disability	106	0.8	104	0.8	16	0.5
Deaf–blindness	28	0.2	27	0.2	12	0.3
Multiple disabilities	1,352	9.9	1,348	9.9	357	10.2
Autism	3,551	25.9	3,546	25.9	805	23.0
Traumatic brain injury	85	0.6	84	0.6	22	0.6
Unknown	165	1.2	163	1.2	20	0.6
TOTAL	13,709	100	13,679	100	3,504	100

* Results for groups with fewer than 11 members are not reported.

Table 8.E.4 CAPA Disability Distributions: Level II

Disability	ELA		Mathematics	
	Frequency	Percent	Frequency	Percent
Mental retardation	2,091	31.5	2,086	31.5
Hard of hearing	35	0.5	35	0.5
Deafness	40	0.6	40	0.6
Speech or language impairment	650	9.8	648	9.8
Visual impairment	36	0.5	36	0.5
Emotional disturbance	23	0.3	23	0.3
Orthopedic impairment	276	4.2	275	4.2
Other health impairment	358	5.4	358	5.4
Specific learning disability	538	8.1	538	8.1
Deaf–blindness*	–	–	–	–
Multiple disabilities	110	1.7	111	1.7
Autism	2,317	34.9	2,306	34.8
Traumatic brain injury	40	0.6	40	0.6
Unknown	125	1.9	124	1.9
TOTAL	6,639	100	6,620	100

Results for groups with fewer than 11 members are not reported.

Table 8.E.5 CAPA Disability Distributions: Level III

Disability	ELA		Mathematics		Science	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Mental retardation	2,643	37.2	2,640	37.2	1,313	38.8
Hard of hearing	61	0.9	61	0.9	28	0.8
Deafness	62	0.9	62	0.9	29	0.9
Speech or language impairment	432	6.1	431	6.1	162	4.8
Visual impairment	42	0.6	42	0.6	28	0.8
Emotional disturbance	39	0.5	39	0.5	21	0.6
Orthopedic impairment	344	4.8	344	4.8	174	5.1
Other health impairment	351	4.9	349	4.9	168	5.0
Specific learning disability	680	9.6	679	9.6	329	9.7
Deaf–blindness*	–	–	–	–	–	–
Multiple disabilities	160	2.3	159	2.2	86	2.5
Autism	2,170	30.5	2,164	30.5	995	29.4
Traumatic brain injury	35	0.5	35	0.5	16	0.5
Unknown	90	1.3	90	1.3	39	1.2
TOTAL	7,109	100	7,095	100	3,388	100

* Results for groups with fewer than 11 members are not reported.

Table 8.E.6 CAPA Disability Distributions: Level IV

Disability	ELA		Mathematics		Science	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Mental retardation	4,434	45.0	4,430	45.0	1,512	47.9
Hard of hearing	53	0.5	53	0.5	24	0.8
Deafness	93	0.9	93	0.9	26	0.8
Speech or language impairment	310	3.1	310	3.2	71	2.3
Visual impairment	58	0.6	58	0.6	22	0.7
Emotional disturbance	91	0.9	90	0.9	26	0.8
Orthopedic impairment	538	5.5	538	5.5	172	5.5
Other health impairment	417	4.2	418	4.2	136	4.3
Specific learning disability	732	7.4	733	7.4	223	7.1
Deaf–blindness*	–	–	–	–	–	–
Multiple disabilities	280	2.8	278	2.8	83	2.6
Autism	2,650	26.9	2,645	26.9	802	25.4
Traumatic brain injury	65	0.7	64	0.7	24	0.8
Unknown	133	1.3	131	1.3	33	1.0
TOTAL	9,854	100	9,841	100	3,154	100

* Results for groups with fewer than 11 members are not reported.

Table 8.E.7 CAPA Disability Distributions: Level V

Disability	ELA		Mathematics		Science	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Mental retardation	4,976	48.7	4,966	48.7	1,605	49.5
Hard of hearing	64	0.6	64	0.6	22	0.7
Deafness	115	1.1	115	1.1	37	1.1
Speech or language impairment	207	2.0	206	2.0	58	1.8
Visual impairment	70	0.7	70	0.7	24	0.7
Emotional disturbance	140	1.4	137	1.3	54	1.7
Orthopedic impairment	526	5.1	524	5.1	165	5.1
Other health impairment	550	5.4	548	5.4	166	5.1
Specific learning disability	809	7.9	811	8.0	271	8.4
Deaf–blindness*	–	–	–	–	–	–
Multiple disabilities	333	3.3	333	3.3	95	2.9
Autism	2,155	21.1	2,151	21.1	678	20.9
Traumatic brain injury	76	0.7	76	0.7	24	0.7
Unknown	195	1.9	194	1.9	46	1.4
TOTAL	10,216	100	10,195	100	3,245	100

* Results for groups with fewer than 11 members are not reported.

Chapter 9: Quality Control Procedures

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

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

Quality Control of Task Development

The task development process for the CAPA is described in detail in Chapter 3, starting on page 16. The next sections highlight elements of the process devoted specifically to the quality control of task development.

Task Specifications

ETS maintains task specifications for the CAPA and has developed an item utilization plan to guide the development of the tasks for each content area. Task writing emphasis is determined in consultation with the CDE. Adherence to the specifications ensures the maintenance of quality and consistency of the task development process.

Task Writers

The tasks for the CAPA are written by task writers that have a thorough understanding of the California content standards. The task writers are carefully screened and selected by senior ETS content staff and approved by the CDE. Only those with strong content and teaching backgrounds who have experience with students who have severe cognitive disabilities are invited to participate in an extensive training program for task writers.

Internal Contractor Reviews

Once tasks have been written, ETS assessment specialists make sure that each task goes through an intensive internal review process. Every step of this process is designed to produce tasks that exceed industry standards for quality. It includes three rounds of content reviews, two rounds of editorial reviews, an internal fairness review, and a high-level review and approval by a content area director. A carefully designed and monitored workflow and detailed checklists help to ensure that all tasks meet the specifications for the process.

Content Review

ETS assessment specialists make sure that the tasks and related materials comply with ETS's written guidelines for clarity, style, accuracy, and appropriateness and with approved task specifications.

The artwork and graphics for the tasks are created during the internal content review period so assessment specialists can evaluate the correctness and appropriateness of the art early in the task development process. ETS selects visuals that are relevant to the task content

and that are easily understood so students do not struggle to determine the purpose or meaning of the questions.

Editorial Review

Another step in the ETS internal review process involves a team of specially trained editors who check questions for clarity, correctness of language, grade-level appropriateness of language, adherence to style guidelines, and conformity to acceptable task-writing practices. The editorial review also includes rounds of copyediting and proofreading. ETS takes pride in the typographical integrity of the tasks presented to our clients and strives for error-free tasks beginning with the initial rounds of review.

Fairness Review

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

Assessment Director Review

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

Assessment Review Panel Review

The ARPs are panels that advise the CDE and ETS on areas related to task development for the CAPA. The ARPs are responsible for reviewing all newly developed tasks for alignment to the California content standards. The ARPs also review the tasks for accuracy of content, clarity of phrasing, and quality. See page 20 in Chapter 3 for additional information on the function of ARPs within the task-review process.

Statewide Pupil Assessment Review Panel Review

The SPAR panel is responsible for reviewing and approving the achievement tests that are to be used statewide for the testing of students in California public schools in grades two through eleven. The SPAR panel representatives ensure that the CAPA tasks conform to the requirements of *EC* Section 60602. See page 22 in Chapter 3 for additional information on the function of the SPAR panel within the task-review process.

Data Review of Field-tested Tasks

ETS field tests newly developed tasks to obtain statistical information about task performance. This information is used to evaluate tasks that are candidates for use in operational test forms. The tasks and task (item) statistics are examined carefully at data review meetings, where content experts discuss tasks that have poor statistics and do not meet the psychometric criteria for task quality. The CDE defines the criteria for acceptable or unacceptable task statistics. These criteria ensure that the task (1) has an appropriate level of difficulty for the target population; (2) discriminates well between examinees that differ in ability; and (3) conforms well to the statistical model underlying the measurement of the intended constructs. The results of analyses for differential item functioning (DIF) are used to make judgments about the appropriateness of items for various subgroups.

The content experts make recommendations about whether to accept or reject each task for inclusion in the California item bank.

Quality Control of the Item Bank

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

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

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

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

- Electronic forms of test content, documentation, and item banks are backed up electronically, with the backup media kept 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.
- Advanced network security measures are used to prevent unauthorized electronic access to the item bank.

Quality Control of Test Form Development

The ETS Assessment Development group is committed to providing the highest quality product to the students of California and has in place a number of quality control (QC) checks to ensure that outcome. During the task development process, there are multiple senior reviews of tasks, including one by the assessment director. Test forms certification is a formal quality control process established as a final checkpoint prior to printing. In it, content, editorial, and senior development staff review test forms for accuracy and clueing issues.

ETS also includes quality checks throughout preparation of the form planners. A form planner specifications document is developed by the test development team lead with input from ETS's item bank and statistics groups; this document is then reviewed by all team members who build forms at a training specific to form planners before the form-building process starts. After trained content team members sign off on a form planner, a representative from the internal QC group reviews each file for accuracy against the

specifications document. Assessment directors review and sign off on form planners prior to processing.

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

Quality Control of Test Materials

Collecting Test Materials

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

All scorable materials are delivered to the Pearson scanning and scoring facilities in Iowa City, Iowa. The nonscorable materials, including *CAPA Examiner's Manuals*, are returned to the Security Processing Department in Pearson's Cedar Rapids, Iowa facility. ETS and Pearson closely monitor the return of materials. The STAR Technical Assistance Center (TAC) at ETS monitors returns and notifies school districts that do not return their materials in a timely manner. STAR TAC contacts the district STAR coordinators and works with them to facilitate the return of the test materials.

Processing Test Materials

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

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

Quality Control of Scanning

Before any STAR documents are scanned, Pearson conducts a complete check of the scanning system. ETS and Pearson create test decks for every test and form. Each test deck consists of approximately 25 answer documents marked to cover response ranges, demographic data, blanks, double marks, and other responses. Fictitious students are created to verify that each marking possibility is processed correctly by the scanning program. The output file generated as a result of this activity is thoroughly checked against each answer document after each stage to verify that the scanner is capturing marks correctly. When the program output is confirmed to match the expected results, a scan program release form is signed and the scan program is placed in the production environment under configuration management.

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

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

Post-scanning Edits

After scanning, there are three opportunities for demographic data to be edited:

- After scanning, by Pearson online editors
- After Pearson's online editing, by district STAR coordinators (demographic edit)
- After paper reporting, by district STAR coordinators

Demographic edits completed by the Pearson editors and by the district STAR coordinator online are included in the data used for the paper reporting and for the technical reports.

Quality Control of Image Editing

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

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

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

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

Quality Control of Answer Document Processing and Scoring

Accountability of Answer Documents

In addition to the quality control checks carried out in scanning and image editing, the following manual quality checks are conducted to verify that the answer documents are correctly attributed to the students, schools, districts, and subgroups:

- Grade counts are compared to the District Master File Sheets.
- Document counts are compared to the School Master File Sheets.
- Document counts are compared to the SGIDs.

Any discrepancies identified in the steps outlined above are followed up by Pearson staff with the school districts for resolution.

Processing of Answer Documents

Prior to processing operational answer sheets and executing subsequent data processing programs, ETS conducts an end-to-end test. As part of this test, ETS prepares approximately 700 test cases covering all tests and many scenarios designed to exercise particular business rule logic. ETS marks answer sheets for those 700 test cases. They are then scanned, scored, and aggregated. The results at various inspection points are checked by psychometricians and Data Quality Services staff. Additionally, a post-scan test file of approximately 50,000 records across the STAR Program is scored and aggregated to test a broader range of scoring and aggregation scenarios. These procedures assure that students and school districts get the correct scores when the actual scoring process is carried out.

Scoring and Reporting Specifications

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

- General Reporting Specifications
- Form Planner Specifications
- Aggregation Rules
- ”What If”. . . List
- Edit Specifications
- Matching Criteria for 10 percent of the CAPA tests that are scored more than once

Each of these documents is explained in detail in Chapter 7, starting on page 46. The scoring specifications are reviewed and revised by the CDE, ETS, and Pearson each year. After a version that all parties endorse is finalized, the CDE issues a formal approval of the scoring and reporting specifications.

Matching Information on CAPA Answer Documents

Answer documents are designed to produce a single complete record for each student. This record includes demographic data and scanned responses for each student; once computed, the scored responses and the total test scores for a student are also merged into the same record. All scores must comply with the ETS scoring specifications.

All STAR answer documents contain uniquely numbered lithocodes that are both scannable and eye-readable. The lithocodes allow all pages of the document to be linked throughout processing, even after the documents have been slit into single sheets for scanning. For those students using more than one score, lithocodes link their demographics and responses within a document, while matching criteria are used to create a single record for all of the student’s documents. The documents are matched within grade using the match criteria approved by the CDE.

Storing Answer Documents

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

Quality Control of Psychometric Processes

Quality Control of Task (Item) Analyses, DIF, and the Scoring Process

The psychometric analyses conducted at ETS undergo comprehensive quality checks by a team of psychometricians and data analysts. Detailed checklists are consulted by members of the team for each of the statistical procedures performed on each CAPA. Quality assurance checks also include a comparison of the current year's statistics to statistics from previous years. The results of preliminary classical task analyses that provide a check on scoring keys are also reviewed by a senior psychometrician. The tasks that are flagged for questionable statistical attributes are sent to test development staff for their review; their comments are reviewed by the psychometricians before tasks are approved to be included in the equating process.

The results of the equating process are reviewed by a psychometric manager in addition to the aforementioned team of psychometricians and data analysts. If the senior psychometrician and the manager reach a consensus that an equating result does not conform to the norm, special binders are prepared for review by senior psychometric advisors at ETS along with several pieces of informative analyses to facilitate the process.

A few additional checks are performed for each process as described below.

Calibrations

During the calibration process, which is described in detail in Chapter 2 starting on page 12, checks are made to ascertain that the correct options for the analyses are selected. Checks are also made on the number of tasks, number of examinees with valid scores, IRT Rasch task difficulty estimates, standard errors for the Rasch task difficulty estimates, and the match of selected statistics to the results on the same statistics obtained during preliminary task analyses. Psychometricians also perform detailed reviews of plots and statistics to investigate if the data fit the model.

Scaling

During the scaling process, checks are made to ensure the following:

- Correct linking tasks are used;
- Stability analysis and subsequent removal of tasks from the linking set (if any) during the scaling evaluation process are implemented according to specification (see details in the "Evaluation of Scaling" section in Chapter 8, on page 82); and
- Scaling constants are correctly applied to transform the new task difficulty estimates on to the item bank scale.

Scoring Tables

Once the equating activities are complete and raw-to-scale scoring tables are generated, the psychometricians carry out quality control checks on each scoring table. Scoring tables are checked to verify the following:

- All raw scores are included in the tables;
- Scale scores increase as raw scores increase;
- The minimum reported scale score is 15 and maximum reported scale score is 60; and
- The cut points for the performance levels are correctly identified.

As a check on the reasonableness of the performance levels, psychometricians compare results from the current year with results from the past year at the cut points and the

percentage of all students in each performance level. After all quality control steps are completed and any differences are resolved, a senior psychometrician checks the scoring tables as the final step in quality control.

Score Verification Process

Pearson utilizes the raw-to-scale scoring tables to assign scale scores for each student. ETS verifies Pearson's scale scores by independently generating the scale scores for students in a small number of school districts and comparing these scores with those generated by Pearson. The selection of districts is based on the availability of data for all schools included in those districts, known as "pilot districts".

Year-to-Year Comparison Analyses

Year-to-year comparison analyses are conducted each year for quality control of the scoring procedure in general and as reasonableness checks for the CAPA results. Year-to-year comparison analyses use over 90 percent of the entire testing population to look at the tendencies and trends for the state as a whole as well as a few large districts. The results of the year-to-year comparison analyses are provided to the CDE and their reasonableness is jointly discussed. Any anomalies in the results are investigated further and scores are released only after explanations that satisfy both the CDE and ETS are obtained.

Offloads to Test Development

The statistics based on classical task analyses and the IRT analyses are obtained at two different times in the testing cycle. The first time, the statistics are obtained on the equating samples to ensure the quality of equating, and then on larger sample sizes to ensure the stability of the statistics that are to be used for future test assembly. Statistics used to generate DIF flags are also obtained from the larger samples and are provided to test development staff in specially designed Excel spreadsheets called "statistical offloads." The offloads are thoroughly checked by the psychometric staff before their release for test development review.

Quality Control of Reporting

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

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

All reports are required to include a single, accurate CDS code, a charter school number (if applicable), a school district name, and a school name. All elements conform to the CDE's official CDS code and naming records. From the start of processing through scoring and reporting, the CDS Master File is used to verify and confirm accurate codes and names. The CDE Master File is provided by the CDE to ETS throughout the year as updates are available.

For students for whom there is more than one answer document, the matching process, as described previously, provides for the creation of individual student records from which reports are created.

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

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

Excluding Student Scores from Summary Reports

ETS provides specifications to the CDE that document when to exclude student scores from summary reports. These specifications include the logic for handling answer documents that, for example, indicate the student was absent, was not tested due to parent/guardian request, or did not complete the test due to illness.

Reference

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

Chapter 10: Historical Comparisons

Base Year Comparisons

Historical comparisons of the CAPA results are routinely performed to identify the trends in examinee performance and test characteristics over time. Such comparisons were performed over a period of the two most recent years of administration—2011 and 2010—and the 2009 base year.

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

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

The CAPA was first administered in 2003. Subsequently, the CAPA has been revised to better link it to the grade-level California content standards. The revised blueprints for the CAPA were approved by the SBE in 2006 for implementation beginning in 2008; new tasks were developed to meet the revised blueprints and field tested.

A standard setting was held in the fall of 2008 to establish new cut scores for the below basic, basic, proficient, and advanced performance levels based on the revised standards for Levels I through V in ELA and mathematics and Levels I and III through V in science. Spring 2009 was the first administration in which test results were reported using the new scales and cut scores for the four performance levels; thus, 2009 became the base year.

Examinee Performance

Table 10.A.1 contains the number of examinees assessed and the means and standard deviations of examinees' scale scores in the base year (2009), in 2010, and in 2011 for each CAPA. As noted in previous chapters, the CAPA reporting scales range from 15 to 60 for all content areas and levels.

CAPA scale scores are used to classify student results into one of five performance levels: far below basic, below basic, basic, proficient, and advanced. The percentages of students qualifying for the proficient and advanced levels are presented in Table 10.A.2 on page 172; please note that this information may differ slightly from information found on the CDE's STAR reporting Web page at http://star.cde.ca.gov due to differing dates on which data were accessed. The goal is for all students to achieve at or above the proficient level by 2014. This goal for all students is consistent with school growth targets for state accountability and the federal requirements under the Elementary and Secondary Education Act.

Table 10.A.3 through Table 10.A.5 show for each CAPA the distribution of scale scores observed in the base year, in 2010, and in 2011. Frequency counts are provided for each scale score interval of 3. A frequency count of "N/A" indicates that there are no obtainable scale scores within that scale score range. For all CAPA, a minimum score of 30 is required for a student to reach the basic level of performance and a minimum score of 35 is required for a student to reach the proficient level of performance.

Test Characteristics

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

Table 10.B.1 and Table 10.B.2 in Appendix 10.B, which starts on page 176, present, respectively, the average item scores and the mean equated IRT *b*-values for the tasks in each CAPA based on the equating samples. The average task scores are affected by both the difficulty of the items and the abilities of the students taking them. The mean equated IRT b-values reflect only average item difficulty. Please note that comparisons of mean *b*-values should only be made within a given test; they should not be compared across test levels or content areas.

The average polyserial correlations for the CAPA are presented in Table 10.B.3. The reliabilities and standard errors of measurement (SEM) expressed in raw score units appear in Table 10.B.4. Like the average item score, polyserial correlations and reliabilities are affected by both item characteristics and student characteristics.

Appendix 10.A—Historical Comparisons Tables

Table 10.A.1 Number of Examinees Tested, Scale Score Means and Standard Deviations of CAPA Across Base Year (2009), 2010, and 2011

		Number of Examinees (valid scores)			Scale Score Mean and Standard Deviation					
					Base		2010		2011	
Content Area	CAPA	Base	2010	2011	Mean	S.D.	Mean	S.D.	Mean	S.D.
English–Language Arts	I	12,531	13,143	13,719	40.84	12.02	40.68	11.33	40.79	10.63
	II	6,587	6,682	6,643	39.24	7.46	38.54	6.25	38.72	5.70
	III	6,614	6,782	7,112	39.12	5.94	39.29	5.83	39.53	6.17
	IV	9,853	9,705	9,858	39.19	7.75	39.15	8.41	39.20	7.44
	V	10,517	10,443	10,217	38.54	6.21	38.73	6.59	38.88	6.44
Mathematics	I	12,484	13,111	13,689	35.11	9.74	35.87	9.30	36.08	8.70
	II	6,569	6,673	6,624	37.60	9.56	37.34	8.37	37.45	7.64
	III	6,602	6,770	7,098	36.58	6.64	36.50	5.80	36.33	4.99
	IV	9,831	9,676	9,845	36.41	8.80	37.15	8.91	37.26	8.11
	V	10,485	10,420	10,196	37.51	8.85	37.52	8.55	37.32	7.63
Science	I	3,296	3,490	3,512	35.59	11.25	36.49	11.13	36.20	10.67
	III	3,267	3,237	3,391	36.24	5.45	36.06	5.02	36.42	5.27
	IV	3,190	3,154	3,155	35.56	5.53	36.24	5.36	36.23	5.72
	V	3,396	3,325	3,245	35.35	5.34	35.69	4.79	35.82	5.08

Table 10.A.2 Percentage of Proficient and Above and Percentage of Advanced Across Base Year (2009), 2010, and 2011

		% Proficient and Above			% Advanced		
Content Area	CAPA	Base	2010	2011	Base	2010	2011
English–Language Arts	I	75%	78%	80%	51%	55%	56%
	II	78%	80%	85%	41%	42%	42%
	III	83%	86%	84%	42%	46%	46%
	IV	77%	72%	79%	37%	43%	35%
	V	80%	79%	79%	42%	45%	48%
Mathematics	I	61%	64%	69%	29%	32%	33%
	II	62%	66%	66%	33%	35%	35%
	III	65%	70%	72%	31%	25%	18%
	IV	60%	68%	67%	31%	33%	32%
	V	67%	68%	68%	34%	37%	34%
Science	I	59%	60%	58%	33%	35%	35%
	III	69%	70%	68%	19%	16%	18%
	IV	58%	71%	65%	15%	18%	20%
	V	61%	64%	68%	17%	19%	20%

Table 10.A.3 Observed Score Distributions of CAPA Across Base Year (2009), 2010, and 2011 for ELA

Observed Score Distributions	Level I			Level II			Level III			Level IV			Level V		
	Base	2010	2011	Base	2010	2011	Base	2010	2011	Base	2010	2011	Base	2010	2011
60	2,230	1,648	1,534	405	78	70	199	86	158	219	150	213	274	221	266
57–59	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
54–56	N/A	456	488	N/A	N/A	N/A	N/A	N/A	N/A	239	189	235	N/A	N/A	N/A
51–53	624	N/A	N/A	N/A	100	91	N/A	118	N/A	N/A	236	N/A	N/A	N/A	N/A
48–50	388	615	580	375	133	156	304	249	300	653	650	664	400	266	346
45–47	299	1,571	1,597	375	491	464	426	693	862	967	1,210	865	517	876	1,029
42–44	1,708	1,555	1,768	795	1,011	978	934	1,191	1,004	1,534	1,764	1,475	1,277	1,634	1,320
39–41	1,784	2,126	2,560	1,090	1,598	1,803	1,341	1,514	1,809	1,911	1,214	2,087	3,097	2,703	2,493
36–38	1,567	1,885	1,883	1,776	1,613	1,487	2,044	1,502	1,360	1,669	1,636	1,795	2,179	2,251	2,190
33–35	1,559	923	953	1,081	836	1,060	891	934	1,030	1,008	845	1,183	1,698	1,167	1,284
30–32	694	599	853	362	507	230	258	255	326	822	787	611	572	824	782
27–29	545	519	428	154	86	129	111	74	115	398	301	340	211	146	250
24–26	140	167	151	89	88	64	45	58	79	83	333	109	113	42	91
21–23	128	135	123	28	36	28	34	49	21	70	65	101	59	106	50
18–20	128	113	156	12	25	24	5	14	16	125	57	37	33	57	32
15–17	737	831	645	45	80	59	22	45	32	155	268	143	87	150	84

Table 10.A.4 Observed Score Distributions of CAPA Across Base Year (2009), 2010 and 2011 for Mathematics

Observed Score Distributions	Level I			Level II			Level III			Level IV			Level V		
	Base	2010	2011	Base	2010	2011	Base	2010	2011	Base	2010	2011	Base	2010	2011
60	603	630	534	417	110	92	134	76	31	269	297	202	767	600	404
57–59	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
54–56	N/A	N/A	N/A	N/A	138	90	N/A	N/A	N/A	N/A	N/A	256	N/A	N/A	N/A
51–53	N/A	N/A	N/A	386	N/A	N/A	N/A	N/A	N/A	391	364	N/A	N/A	N/A	N/A
48–50	237	297	245	N/A	208	143	230	95	47	295	383	296	N/A	381	269
45–47	382	259	247	338	887	613	N/A	126	146	687	371	656	499	369	316
42–44	934	1,111	1,235	682	659	1,071	762	625	497	689	1,344	1,246	1,104	997	1,158
39–41	1,465	1,963	2,286	886	958	941	1,274	1,164	1,384	1,436	1,519	1,585	1,804	1,938	1,849
36–38	2,775	3,252	3,636	1,049	1,172	1,052	1,579	1,979	2,550	1,687	1,611	1,957	2,475	2,356	2,439
33–35	2,628	2,568	2,525	1,053	828	1,109	1,105	1,576	1,582	1,229	1,233	1,087	1,524	1,446	1,745
30–32	1,053	946	973	658	729	803	837	645	413	1,319	644	1,102	918	917	1,022
27–29	407	595	609	547	390	317	320	220	198	888	772	777	473	601	286
24–26	492	216	174	137	259	97	200	133	81	286	562	219	278	245	434
21–23	174	171	159	209	127	142	39	21	58	257	272	190	321	238	48
18–20	177	162	161	34	32	29	33	14	18	75	64	54	61	90	37
15–17	1,157	941	905	173	176	125	89	96	93	323	240	218	261	242	189

Table 10.A.5 Observed Score Distributions of CAPA Across Base Year (2009), 2010, and 2011 for Science

Observed Score Distributions	Level I			Level III			Level IV			Level V		
	Base	2010	2011	Base	2010	2011	Base	2010	2011	Base	2010	2011
60	280	346	293	69	48	69	46	41	61	33	27	33
57–59	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
54–56	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
51–53	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
48–50	81	113	90	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
45–47	69	85	73	105	59	N/A	44	68	71	46	33	46
42–44	267	264	338	122	91	228	157	203	225	129	102	167
39–41	394	427	441	493	482	531	393	398	431	373	475	413
36–38	588	668	706	934	1,220	1,154	1,010	1,144	858	1,288	1,266	1,121
33–35	611	598	594	1,093	804	941	864	791	916	874	922	903
30–32	271	330	299	268	352	300	420	310	382	332	306	353
27–29	108	130	164	104	107	92	155	77	106	196	105	133
24–26	207	144	125	29	23	34	36	60	53	36	23	23
21–23	N/A	40	49	20	23	16	10	19	15	25	15	16
18–20	49	43	43	10	6	11	19	8	13	14	11	4
15–17	371	302	297	20	22	15	36	35	24	50	40	33

Appendix 10.B—Historical Comparisons Tables

Table 10.B.1 Average Item Score of CAPA Operational Test Items Across Base Year (2009), 2010, and 2011

Content Area	CAPA	Average Item Score		
		Base	2010	2011
English–Language Arts	I	3.37	3.18	3.21
	II	2.91	2.43	2.51
	III	2.91	2.61	2.86
	IV	2.51	2.34	2.54
	V	2.73	2.54	2.74
Mathematics	I	2.70	2.80	2.85
	II	2.70	2.57	2.55
	III	2.70	2.48	2.32
	IV	2.37	2.57	2.47
	V	2.76	2.62	2.51
Science	I	2.75	2.98	2.91
	III	2.71	2.48	2.69
	IV	2.47	2.57	2.69
	V	2.47	2.44	2.58

Table 10.B.2 Mean IRT *b*-values for Operational Test Items Across Base Year (2009), 2010, and 2011

Content Area	CAPA	Mean IRT *b*-value		
		Base	2010	2011
English–Language Arts	I	–0.74	–0.63	–0.61
	II	–1.54	–0.91	–0.99
	III	–1.52	–0.95	–1.32
	IV	–0.93	–0.79	–1.05
	V	–1.19	–0.89	–1.16
Mathematics	I	–0.29	–0.23	–0.21
	II	–1.18	–1.04	–1.02
	III	–1.29	–0.95	–0.87
	IV	–0.85	–1.02	–0.86
	V	–1.21	–1.05	–0.98
Science	I	–0.23	–0.37	–0.33
	III	–1.29	–1.01	–1.29
	IV	–0.95	–0.98	–1.19
	V	–0.54	–0.39	–0.52

Table 10.B.3 Mean Polyserial Correlation of CAPA Operational Test Items Across Base Year (2009), 2010, and 2011

Content Area	CAPA	Mean Polyserial Correlation		
		Base	2010	2011
English–Language Arts	I	0.81	0.81	0.79
	II	0.75	0.75	0.72
	III	0.75	0.78	0.78
	IV	0.78	0.80	0.77
	V	0.79	0.81	0.80
Mathematics	I	0.79	0.77	0.75
	II	0.78	0.77	0.73
	III	0.76	0.73	0.68
	IV	0.79	0.79	0.75
	V	0.78	0.78	0.76
Science	I	0.82	0.81	0.80
	III	0.75	0.74	0.73
	IV	0.75	0.73	0.73
	V	0.78	0.74	0.76

Table 10.B.4 Score Reliabilities and SEM of CAPA Across Base Year (2009), 2010, and 2011

Content Area	CAPA	Reliability			SEM		
		Base	2010	2011	Base	2010	2011
English–Language Arts	I	0.91	0.90	0.88	3.67	3.80	3.86
	II	0.84	0.86	0.82	2.49	2.51	2.59
	III	0.86	0.88	0.88	2.26	2.26	2.19
	IV	0.88	0.90	0.86	2.50	2.32	2.48
	V	0.89	0.91	0.89	2.35	2.27	2.19
Mathematics	I	0.87	0.86	0.84	4.00	4.21	4.27
	II	0.88	0.86	0.82	2.58	2.64	2.64
	III	0.87	0.83	0.77	2.54	2.59	2.63
	IV	0.88	0.88	0.85	2.62	2.57	2.64
	V	0.87	0.88	0.85	2.70	2.62	2.73
Science	I	0.91	0.89	0.89	3.76	3.85	3.90
	III	0.85	0.85	0.84	2.43	2.51	2.49
	IV	0.85	0.84	0.84	2.46	2.47	2.32
	V	0.87	0.84	0.86	2.30	2.35	2.26