

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



**California Alternate Performance  
Assessment  
Technical Report  
Spring 2014 Administration**

**Updated Final Version Submitted February 13, 2015  
Educational Testing Service  
Contract No. 5417**



# Table of Contents

---

Acronyms and Initialisms Used in the <i>CAPA Technical Report</i> .....	vii
<b>Chapter 1: Introduction</b> .....	<b>1</b>
<b>Background</b> .....	<b>1</b>
<b>Test Purpose</b> .....	<b>1</b>
<b>Test Content</b> .....	<b>2</b>
<b>Intended Population</b> .....	<b>2</b>
<b>Intended Use and Purpose of Test Scores</b> .....	<b>3</b>
<b>Testing Window</b> .....	<b>3</b>
<b>Significant CAASPP Developments in 2014</b> .....	<b>3</b>
Renamed the Program.....	3
Change in the Date for Students in Ungraded Programs .....	4
Pre-equated All Results .....	4
Reduced the Number of Test Versions .....	4
Suspended Reporting of Adequate Yearly Progress and the Academic Performance Index .....	4
<b>Limitations of the Assessment</b> .....	<b>4</b>
Score Interpretation .....	4
Out-of-Level Testing .....	4
Score Comparison .....	4
<b>Groups and Organizations Involved with the CAASPP Assessment System</b> .....	<b>5</b>
State Board of Education .....	5
California Department of Education .....	5
Contractors .....	5
<b>Overview of the Technical Report</b> .....	<b>6</b>
<b>References</b> .....	<b>7</b>
<b>Chapter 2: An Overview of CAPA Processes</b> .....	<b>8</b>
<b>Task (Item) Development</b> .....	<b>8</b>
Task Formats .....	8
Task (Item) Specifications.....	8
Item Banking.....	8
Task Refresh Rate .....	9
<b>Test Assembly</b> .....	<b>9</b>
Test Length .....	9
Test Blueprints .....	9
Content Rules and Task Selection.....	9
Psychometric Criteria.....	10
<b>Test Administration</b> .....	<b>10</b>
Test Security and Confidentiality.....	10
Procedures to Maintain Standardization .....	11
<b>Universal Tools, Designated Supports, and Accommodations</b> .....	<b>11</b>
<b>Scores</b> .....	<b>12</b>
Aggregation Procedures .....	12
<b>Equating</b> .....	<b>13</b>
Post-Equating .....	13
Calibration.....	13
Scaling .....	13
Linear Transformation.....	14
<b>References</b> .....	<b>16</b>
<b>Chapter 3: Task (Item) Development</b> .....	<b>17</b>
<b>Rules for Task Development</b> .....	<b>17</b>
Task Specifications .....	17
Expected Task Ratio.....	18
<b>Selection of Task Writers</b> .....	<b>18</b>
Criteria for Selecting Task Writers .....	18
<b>Task (Item) Review Process</b> .....	<b>19</b>
Contractor Review .....	19
Content Expert Reviews .....	20
Statewide Pupil Assessment Review Panel .....	22
<b>Field Testing</b> .....	<b>22</b>
Stand-alone Field Testing .....	23

Embedded Field-test Tasks .....	23
<b>CDE Data Review</b> .....	<b>23</b>
<b>Item Banking</b> .....	<b>23</b>
<b>References</b> .....	<b>25</b>
<b>Chapter 4: Test Assembly</b> .....	<b>26</b>
<b>Test Length</b> .....	<b>26</b>
<b>Rules for Task Selection</b> .....	<b>26</b>
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
<b>Chapter 5: Test Administration</b> .....	<b>29</b>
<b>Test Security and Confidentiality</b> .....	<b>29</b>
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
<b>Procedures to Maintain Standardization</b> .....	<b>34</b>
Test Administrators.....	34
CAPA Examiner's Manual.....	35
LEA CAASPP and Test Site Coordinator Manual.....	36
Test Management System Manuals.....	36
<b>Universal Tools, Designated Supports, and Accommodations for Students with Disabilities</b> .....	<b>36</b>
Identification.....	36
Adaptations.....	37
Scoring.....	37
<b>Testing Incidents</b> .....	<b>37</b>
Social Media Security Breaches.....	37
<b>Testing Improprieties</b> .....	<b>38</b>
<b>References</b> .....	<b>39</b>
<b>Chapter 6: Performance Standards</b> .....	<b>40</b>
<b>Background</b> .....	<b>40</b>
<b>Standard-Setting Procedure</b> .....	<b>40</b>
Development of Competencies Lists.....	41
<b>Standard-Setting Methodology</b> .....	<b>42</b>
Performance Profile Method.....	42
<b>Results</b> .....	<b>42</b>
<b>References</b> .....	<b>44</b>
<b>Chapter 7: Scoring and Reporting</b> .....	<b>45</b>
<b>Procedures for Maintaining and Retrieving Individual Scores</b> .....	<b>45</b>
Scoring and Reporting Specifications.....	46
Scanning and Scoring.....	46
<b>Types of Scores</b> .....	<b>47</b>
Raw Score.....	47
Scale Score.....	47
Performance Levels.....	47
<b>Score Verification Procedures</b> .....	<b>47</b>
Monitoring and Quality Control of Scoring.....	47
Score Verification Process.....	48
<b>Overview of Score Aggregation Procedures</b> .....	<b>48</b>

Individual Scores.....	48
<b>Reports Produced and Scores for Each Report .....</b>	<b>52</b>
Types of Score Reports .....	52
Score Report Contents .....	53
Score Report Applications.....	53
<b>Criteria for Interpreting Test Scores .....</b>	<b>54</b>
<b>Criteria for Interpreting Score Reports.....</b>	<b>54</b>
<b>References .....</b>	<b>55</b>
<b>Appendix 7.A—Scale Score Distribution Tables .....</b>	<b>56</b>
<b>Appendix 7.B—Demographic Summaries.....</b>	<b>58</b>
<b>Appendix 7.C—Types of Score Reports.....</b>	<b>64</b>
<b>Chapter 8: Analyses.....</b>	<b>67</b>
<b>Samples Used for the Analyses .....</b>	<b>67</b>
<b>Classical Analyses .....</b>	<b>68</b>
Average Item Score .....	68
Polyserial Correlation of the Task Score with the Total Test Score .....	68
<b>Reliability Analyses.....</b>	<b>70</b>
Subgroup Reliabilities and SEMs.....	71
Conditional Standard Errors of Measurement.....	71
<b>Decision Classification Analyses .....</b>	<b>72</b>
<b>Validity Evidence.....</b>	<b>73</b>
Purposes of the CAPA .....	74
The Constructs to Be Measured .....	74
Interpretations and Uses of the Scores Generated .....	74
Intended Test Population(s).....	75
Validity Evidence Collected.....	75
Evidence Based on Response Processes .....	78
Evidence of Interrater Agreement .....	78
Evidence Based on Internal Structure.....	78
Evidence Based on Consequences of Testing.....	79
<b>IRT Analyses.....</b>	<b>79</b>
Post-Equating .....	79
Pre-Equating.....	80
Summaries of Scaled IRT <i>b</i> -values.....	80
Equating Results.....	80
<b>Differential Item Functioning Analyses .....</b>	<b>81</b>
<b>References .....</b>	<b>84</b>
<b>Appendix 8.A—Classical Analyses: Task Statistics .....</b>	<b>86</b>
<b>Appendix 8.B—Reliability Analyses .....</b>	<b>96</b>
<b>Appendix 8.C—Validity Analyses .....</b>	<b>110</b>
<b>Appendix 8.D—IRT Analyses .....</b>	<b>127</b>
<b>Appendix 8.E—Disability Distributions .....</b>	<b>141</b>
<b>Chapter 9: Quality Control Procedures.....</b>	<b>144</b>
<b>Quality Control of Task Development.....</b>	<b>144</b>
Task Specifications .....	144
Task Writers.....	144
Internal Contractor Reviews.....	144
Assessment Review Panel Review .....	145
Statewide Pupil Assessment Review Panel Review .....	145
Data Review of Field-tested Tasks .....	145
<b>Quality Control of the Item Bank.....</b>	<b>146</b>
<b>Quality Control of Test Form Development .....</b>	<b>146</b>
<b>Quality Control of Test Materials .....</b>	<b>147</b>
Collecting Test Materials.....	147
Processing Test Materials.....	147
<b>Quality Control of Scanning.....</b>	<b>147</b>
<b>Quality Control of Image Editing.....</b>	<b>148</b>
<b>Quality Control of Answer Document Processing and Scoring.....</b>	<b>148</b>
Accountability of Answer Documents .....	148
Processing of Answer Documents .....	148
Scoring and Reporting Specifications .....	149
Storing Answer Documents.....	149

<b>Quality Control of Psychometric Processes</b> .....	<b>149</b>
Quality Control of Task (Item) Analyses and the Scoring Process.....	149
Score Verification Process.....	150
Year-to-Year Comparison Analyses.....	150
Offloads to Test Development.....	151
<b>Quality Control of Reporting</b> .....	<b>151</b>
Electronic Reporting.....	151
Excluding Student Scores from Summary Reports.....	152
<b>Reference</b> .....	<b>153</b>
<b>Chapter 10: Historical Comparisons</b> .....	<b>154</b>
<b>Base Year Comparisons</b> .....	<b>154</b>
<b>Examinee Performance</b> .....	<b>154</b>
<b>Test Characteristics</b> .....	<b>155</b>
<b>Appendix 10.A—Historical Comparisons Tables, Examinee Performance</b> .....	<b>156</b>
<b>Appendix 10.B—Historical Comparisons Tables, Test Characteristics</b> .....	<b>160</b>

**Tables**

Table 1.1 Description of the CAPA Assessment Levels.....	2
Table 2.1 CAPA Items and Estimated Time Chart.....	9
Table 2.2 Scale Score Ranges for Performance Levels.....	14
Table 4.1 Statistical Targets for CAPA Test Assembly.....	27
Table 4.2 Summary of 2013 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.....	50
Table 7.6 Subgroup Definitions.....	51
Table 7.7 Types of CAPA Reports.....	52
Table 7.A.1 Scale Score Frequency Distributions: ELA, Levels I–V.....	56
Table 7.A.2 Scale Score Frequency Distributions: Mathematics, Levels I–V.....	56
Table 7.A.3 Scale Score Frequency Distributions: Science, Levels I–V.....	57
Table 7.B.1 Demographic Summary for ELA, All Examinees.....	58
Table 7.B.2 Demographic Summary for Mathematics, All Examinees.....	60
Table 7.B.3 Demographic Summary for Science, All Examinees.....	62
Table 7.C.1 Score Reports Reflecting CAPA Results.....	64
Table 8.1 CAPA Raw Score Means and Standard Deviations: Total Population for 2013 and 2014.....	68
Table 8.2 Average Item Score and Polyserial Correlation.....	69
Table 8.3 Reliabilities and SEMs for the CAPA.....	71
Table 8.4 CAPA Content-area Correlations for CAPA Levels.....	78
Table 8.5 DIF Flags Based on the ETS DIF Classification Scheme.....	82
Table 8.6 Subgroup Classification for DIF Analyses.....	83
Table 8.A.1 AIS and Polyserial Correlation: Level I, ELA—Current Year (2014) and Original Year of Administration (2013).....	86
Table 8.A.2 AIS and Polyserial Correlation: Level II, ELA—Current Year (2014) and Original Year of Administration (2013).....	86
Table 8.A.3 AIS and Polyserial Correlation: Level III, ELA—Current Year (2014) and Original Year of Administration (2013).....	87
Table 8.A.4 AIS and Polyserial Correlation: Level IV, ELA—Current Year (2014) and Original Year of Administration (2013).....	87
Table 8.A.5 AIS and Polyserial Correlation: Level V, ELA—Current Year (2014) and Original Year of Administration (2013).....	88
Table 8.A.6 AIS and Polyserial Correlation: Level I, Mathematics—Current Year (2014) and Original Year of Administration (2013).....	88
Table 8.A.7 AIS and Polyserial Correlation: Level II, Mathematics—Current Year (2014) and Original Year of Administration (2013).....	89
Table 8.A.8 AIS and Polyserial Correlation: Level III, Mathematics—Current Year (2014) and Original Year of Administration (2013).....	89
Table 8.A.9 AIS and Polyserial Correlation: Level IV, Mathematics—Current Year (2014) and Original Year of Administration (2013).....	90
Table 8.A.10 AIS and Polyserial Correlation: Level V, Mathematics—Current Year (2014) and Original Year of Administration (2013).....	90
Table 8.A.11 AIS and Polyserial Correlation: Level I, Science—Current Year (2014) and Original Year of Administration (2013).....	91

Table 8.A.12 AIS and Polyserial Correlation: Level III, Science—Current Year (2014) and Original Year of Administration (2013).....	91
Table 8.A.13 AIS and Polyserial Correlation: Level IV, Science—Current Year (2014) and Original Year of Administration (2013).....	92
Table 8.A.14 AIS and Polyserial Correlation: Level V, Science—Current Year (2014) and Original Year of Administration (2013).....	92
Table 8.A.15 Frequency of Operational Task Scores: ELA.....	93
Table 8.A.16 Frequency of Operational Task Scores: Mathematics.....	94
Table 8.A.17 Frequency of Operational Task Scores: Science.....	95
Table 8.B.1 Reliabilities and SEMs by Gender.....	96
Table 8.B.2 Reliabilities and SEMs by Primary Ethnicity.....	97
Table 8.B.3 Reliabilities and SEMs by Primary Ethnicity for Economically Disadvantaged.....	98
Table 8.B.4 Reliabilities and SEMs by Primary Ethnicity for Not Economically Disadvantaged.....	99
Table 8.B.5 Reliabilities and SEMs by Primary Ethnicity for Unknown Economic Status.....	100
Table 8.B.6 Reliabilities and SEMs by Disability.....	101
Table 8.B.7 Decision Accuracy and Decision Consistency: Level I, ELA.....	103
Table 8.B.8 Decision Accuracy and Decision Consistency: Level I, Mathematics.....	103
Table 8.B.9 Decision Accuracy and Decision Consistency: Level I, Science.....	104
Table 8.B.10 Decision Accuracy and Decision Consistency: Level II, ELA.....	104
Table 8.B.11 Decision Accuracy and Decision Consistency: Level II, Mathematics.....	105
Table 8.B.12 Decision Accuracy and Decision Consistency: Level III, ELA.....	105
Table 8.B.13 Decision Accuracy and Decision Consistency: Level III, Mathematics.....	106
Table 8.B.14 Decision Accuracy and Decision Consistency: Level III, Science.....	106
Table 8.B.15 Decision Accuracy and Decision Consistency: Level IV, ELA.....	107
Table 8.B.16 Decision Accuracy and Decision Consistency: Level IV, Mathematics.....	107
Table 8.B.17 Decision Accuracy and Decision Consistency: Level IV, Science.....	108
Table 8.B.18 Decision Accuracy and Decision Consistency: Level V, ELA.....	108
Table 8.B.19 Decision Accuracy and Decision Consistency: Level V, Mathematics.....	109
Table 8.B.20 Decision Accuracy and Decision Consistency: Level V, Science.....	109
Table 8.C.1 CAPA Content Area Correlations by Gender: Level I.....	110
Table 8.C.2 CAPA Content Area Correlations by Gender: Level II.....	110
Table 8.C.3 CAPA Content Area Correlations by Gender: Level III.....	110
Table 8.C.4 CAPA Content Area Correlations by Gender: Level IV.....	110
Table 8.C.5 CAPA Content Area Correlations by Gender: Level V.....	110
Table 8.C.6 CAPA Content Area Correlations by Ethnicity: Level I.....	111
Table 8.C.7 CAPA Content Area Correlations by Ethnicity: Level II.....	111
Table 8.C.8 CAPA Content Area Correlations by Ethnicity: Level III.....	111
Table 8.C.9 CAPA Content Area Correlations by Ethnicity: Level IV.....	111
Table 8.C.10 CAPA Content Area Correlations by Ethnicity: Level V.....	112
Table 8.C.11 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level I.....	112
Table 8.C.12 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level II.....	112
Table 8.C.13 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level III.....	112
Table 8.C.14 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level IV.....	113
Table 8.C.15 CAPA Content Area Correlations by Ethnicity for Economically Disadvantaged: Level V.....	113
Table 8.C.16 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level I.....	113
Table 8.C.17 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level II.....	113
Table 8.C.18 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level III.....	114
Table 8.C.19 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level IV.....	114
Table 8.C.20 CAPA Content Area Correlations by Ethnicity for Not Economically Disadvantaged: Level V.....	114
Table 8.C.21 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level I.....	114
Table 8.C.22 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level II.....	115
Table 8.C.23 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level III.....	115
Table 8.C.24 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level IV.....	115
Table 8.C.25 CAPA Content Area Correlations by Ethnicity for Unknown Economic Status: Level V.....	115
Table 8.C.26 CAPA Content Area Correlations by Economic Status: Level I.....	116
Table 8.C.27 CAPA Content Area Correlations by Economic Status: Level II.....	116
Table 8.C.28 CAPA Content Area Correlations by Economic Status: Level III.....	116
Table 8.C.29 CAPA Content Area Correlations by Economic Status: Level IV.....	116
Table 8.C.30 CAPA Content Area Correlations by Economic Status: Level V.....	116
Table 8.C.31 CAPA Content Area Correlations by Disability: Level I.....	117
Table 8.C.32 CAPA Content Area Correlations by Disability: Level II.....	118
Table 8.C.33 CAPA Content Area Correlations by Disability: Level III.....	119
Table 8.C.34 CAPA Content Area Correlations by Disability: Level IV.....	120
Table 8.C.35 CAPA Content Area Correlations by Disability: Level V.....	121
Table 8.C.36 Interrater Agreement Analyses for Operational Tasks: Level I.....	122

Table 8.C.37 Interrater Agreement Analyses for Operational Tasks: Level II .....	123
Table 8.C.38 Interrater Agreement Analyses for Operational Tasks: Level III .....	124
Table 8.C.39 Interrater Agreement Analyses for Operational Tasks: Level IV .....	125
Table 8.C.40 Interrater Agreement Analyses for Operational Tasks: Level V .....	126
Table 8.D.1 Score Conversions: Level I, ELA .....	127
Table 8.D.2 Score Conversions: Level II, ELA .....	128
Table 8.D.3 Score Conversions: Level III, ELA .....	129
Table 8.D.4 Score Conversions: Level IV, ELA .....	130
Table 8.D.5 Score Conversions: Level V, ELA .....	131
Table 8.D.6 Score Conversions: Level I, Mathematics .....	132
Table 8.D.7 Score Conversions: Level II, Mathematics .....	133
Table 8.D.8 Score Conversions: Level III, Mathematics .....	134
Table 8.D.9 Score Conversions: Level IV, Mathematics .....	135
Table 8.D.10 Score Conversions: Level V, Mathematics .....	136
Table 8.D.11 Score Conversions: Level I, Science .....	137
Table 8.D.12 Score Conversions: Level III, Science .....	138
Table 8.D.13 Score Conversions: Level IV, Science .....	139
Table 8.D.14 Score Conversions: Level V, Science .....	140
Table 8.E.1 CAPA Primary Disability Distributions: Level I .....	141
Table 8.E.2 CAPA Primary Disability Distributions: Level II .....	141
Table 8.E.3 CAPA Primary Disability Distributions: Level III .....	142
Table 8.E.4 CAPA Primary Disability Distributions: Level IV .....	142
Table 8.E.5 CAPA Primary Disability Distributions: Level V .....	143
Table 10.A.1 Number of Examinees Tested, Scale Score Means, and Standard Deviations of CAPA Across Base Year (2009), 2012, 2013, and 2014 .....	156
Table 10.A.2 Percentage of Proficient and Above and Percentage of Advanced Across Base Year (2009), 2012, 2013, and 2014 .....	156
Table 10.A.3 Observed Score Distributions of CAPA Across Base Year (2009), 2012, 2013, and 2014 for ELA .....	157
Table 10.A.4 Observed Score Distributions of CAPA Across Base Year (2009), 2012, 2013, and 2014 for Mathematics .....	158
Table 10.A.5 Observed Score Distributions of CAPA Across Base Year (2009), 2012, 2013, and 2014 for Science .....	159
Table 10.B.1 Average Item Score of CAPA Operational Test Tasks Across Base Year (2009), 2012, 2013, and 2014 .....	160
Table 10.B.2 Mean IRT <i>b</i> -values for Operational Test Tasks Across Base Year (2009), 2012, 2013, and 2014 .....	160
Table 10.B.3 Mean Polyserial Correlation of CAPA Operational Test Tasks Across Base Year (2009), 2012, 2013, and 2014 .....	161
Table 10.B.4 Score Reliabilities and SEM of CAPA Across Base Year (2009), 2012, 2013, and 2014 .....	161

## Figures

Figure 3.1 The ETS Item Development Process for the CAASPP System .....	17
Figure 8.1 Decision Accuracy for Achieving a Performance Level .....	73
Figure 8.2 Decision Consistency for Achieving a Performance Level .....	73

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

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



# 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.

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

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

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

## Test Purpose

The CAPA program is designed to show how well students with significant cognitive disabilities are performing with respect to California’s content standards for English–language arts (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; the CAPA links directly to them 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 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

The CAPA for ELA, mathematics, and science are administered to students in one of five levels.

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 2014 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
<b>Grades</b>	2–11	2 and 3	4 and 5	6–8	9–11
<b>Content Area</b>	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
<b>Age Ranges for Ungraded Programs *</b>	7–16	7 & 8	9 & 10	11–13	14–16

\* For students in ungraded programs and whose IEP teams designate that they take the CAPA, their grade is determined by subtracting five from their chronological age on October 1, 2013.

## Intended Population

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 non-embedded accessibility supports 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 CAASPP System. 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 CAASPP System are used for three primary purposes, described as follows (excerpted from the *EC* Section 60602 Web page at <http://www.leginfo.ca.gov/cgi-bin/displaycode?section=edc&group=60001-61000&file=60600-60603>):

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

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

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

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

## Testing Window

The CAPA are administered within a 25-day window, which begins 12 days before and ends 12 days after the day on which 85 percent of the instructional year is completed.

The CAPA 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 LEA’s testing window (*California Code of Regulations [CCR], Title 5, Education, Division 1, Chapter 2, Subchapter 3.75, Article 2, § 855[a][2]*).

## Significant CAASPP Developments in 2014

### Renamed the Program

The paper-pencil tests administered in 2014 are a component of the California Assessment of Student Performance and Progress (CAASPP) Assessment System.

## **Change in the Date for Students in Ungraded Programs**

The date used for determining the testing grade of a student in an ungraded program has changed; for 2013–14, it is October 1, 2013 (*EC* Section 48000[a][2]).

## **Pre-equated All Results**

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

## **Reduced the Number of Test Versions**

Only one version of the CAPA was administered in 2014. All LEAs used the same version.

## **Suspended Reporting of Adequate Yearly Progress and the Academic Performance Index**

The Adequate Yearly Progress (AYP) report submitted to the U.S. Department of Education in 2014 includes CAPA results for ELA and mathematics. Reporting of Academic Performance Index (API) data has been suspended.

## **Limitations of the Assessment**

### **Score Interpretation**

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

### **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 CAASPP System; all students are required to take the test for the grade in which they are enrolled. LEAs are advised to review all IEPs to ensure that any provision for testing below a student's grade level has been removed.

### **Score Comparison**

When comparing 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 2014 CAPA Level II (Mathematics) test. Similarly, it is appropriate to compare scores obtained on the 2012 CAPA Level IV (ELA) test with those obtained on the CAPA Level IV (ELA) test administered in 2014. 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, results from tests administered after 2009 cannot meaningfully be compared to results obtained in previous years.

# Groups and Organizations Involved with the CAASPP Assessment System

## State Board of Education

The SBE is responsible for assuring the compliance with programs that meet the requirement of the federal Elementary and Secondary Education Act (ESEA) and the state's Public School Accountability Act (PSAA) and for reporting CAPA results in grade ten ELA and mathematics in terms of the AYP and, in previous years, the API; these measure the academic performance and growth of schools on a variety of academic metrics. (Data are not reported for either of these measures in 2014 for the CSTs or CMA.)

## California Department of Education

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

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

## Contractors

### Educational Testing Service

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

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

### Pearson

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

- Production of all scannable test materials;
- Packaging, distribution, and collection of testing materials to LEAs and independently testing charter schools;

- Scanning and scoring of all responses, including performance scoring of the writing responses; and
- Production of all score reports and data files of test results.

## Overview of the Technical Report

This technical report addresses the characteristics of the CAPA administered in spring 2014. The technical report contains nine additional chapters as follows:

- Chapter 2 presents a conceptual overview of processes involved in a testing cycle for a CAPA form. This includes test construction, test administration, generation of test scores, and dissemination of score reports. Information about the distributions of scores aggregated by subgroups based on demographics and the use of special services is included, as are the references to various chapters that detail the processes briefly discussed in this chapter.
- Chapter 3 describes the procedures followed during the development of valid CAPA tasks before the 2014 administration—in 2014, intact test forms (form 1 of each CAPA level) from the 2013 administrations were reused and there was no new item development. The chapter also 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 forms reused in 2014.
- Chapter 5 presents the processes involved in the actual administration of the 2014 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 previously 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 and includes a discussion of quick-turnaround reporting.
- Chapter 8 summarizes the results of the task (item)-level analyses performed during the spring 2014 administration of the tests. These include the classical item analyses, the reliability analyses that include assessments of test reliability and the consistency and accuracy of the CAPA performance-level classifications, and the procedures designed to ensure the validity of CAPA score uses and interpretations.
- 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 three 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.

## References

- California Code of Regulations, Title 5, Education, Division 1, Chapter 2, Subchapter 3.75, Article 2, § 855.* Retrieved from <http://www.cde.ca.gov/re/lr/rr/caaspp.asp>
- California Department of Education. (2013). *STAR Program information packet for school district and school staff* (p. 15). Sacramento, CA.
- California Department of Education, EdSource, & the Fiscal Crisis Management Assistance Team. (2014). *Fiscal, demographic, and performance data on California's K–12 schools*. Sacramento, CA: Ed-Data. Retrieved from [http://www.ed-data.k12.ca.us/App\\_Resx/EdDataClassic/fsTwoPanel.aspx?#!bottom=\\_layouts/EdDataClassic/profile.asp?Tab=1&level=04&reportNumber=16](http://www.ed-data.k12.ca.us/App_Resx/EdDataClassic/fsTwoPanel.aspx?#!bottom=_layouts/EdDataClassic/profile.asp?Tab=1&level=04&reportNumber=16)

## **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 implement each of those processes. In 2014, due to the use of intact test forms from the 2013 CAASPP administration, neither test development nor equating activities occurred.

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 were developed to measure California content standards and designed to conform to principles of task writing defined by ETS (ETS, 2002). ETS maintained and updated a task specifications document, otherwise known as “task writer guidelines,” for each CAPA and used an item utilization plan to guide the development of the tasks for each content area. Task writing emphasis was determined in consultation with the CDE.

The task specifications described the characteristics of the tasks that should be written to measure each content standard; tasks of the same type should consistently measure the content standards in the same way. To achieve this, the task specifications provided detailed information to task writers who developed tasks for the CAPA.

The tasks selected for each CAPA underwent 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 17.

#### **Item Banking**

Before newly developed tasks were placed in the item bank, ETS prepared them 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 was complete, the tasks were placed in the item bank along with the associated information obtained at the review sessions. Tasks that were accepted by the content experts were updated to a “field-test ready” status. ETS then delivered the tasks to the CDE by means of a delivery of the California electronic item bank. Tasks were subsequently field-tested to obtain information about task performance and task (item) statistics that could be used to assemble operational forms.

The CDE then reviewed those tasks with their statistical data flagged to determine whether they should be used operationally (see page 23 for more information about the CDE's data review). Any additional updates to task content and statistics were based on data collected

from the operational use of the tasks. However, only the latest content of the task is retained in the bank at any time, along with the administration data from every administration that has included the task.

Further details on item banking are presented on page 23 in Chapter 3.

### Task Refresh Rate

Prior to form reuse in the 2014 administration, the item utilization plan required that each year, 25 percent of tasks on an operational form were refreshed (replaced); these tasks remained 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 necessary within the LEA's selected testing window.

**Table 2.1 CAPA Items and Estimated Time Chart**

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

### Test Blueprints

ETS selected all CAPA tasks to conform to the SBE-approved California content standards and test blueprints. The revised blueprints for the CAPA were approved by the SBE in 2006 for implementation beginning in 2008. The test blueprints for the CAPA are linked on the CDE CAASPP Science Assessments Web page at <http://www.cde.ca.gov/ta/tg/ca/caasppscience.asp>.

### Content Rules and Task Selection

Intact test forms from last year's CAPA administration were reused during the 2014 administration. Prior to the 2014 administration, test developers follow a number of rules when developing a new test form for a given CAPA level and content area. First and foremost, they selected tasks that met the blueprint for that level and content area. Using the electronic item bank, assessment specialists began by identifying a number of linking tasks. These were tasks that appeared in previous operational test administrations and were then used to equate the subsequent (new) test forms. After the linking tasks were approved, assessment specialists populated the rest of the test form.

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

Another consideration was the difficulty of each task. Test developers strived to ensure that there were some easy and some hard tasks and that there were 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 assessed the projected test characteristics during the preliminary review of the assembled forms. The statistical targets used to develop the 2013 forms and the projected characteristics of the assembled forms are presented starting from page 27 in Chapter 4.

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

### **Test Administration**

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

### **Test Security and Confidentiality**

All tests within the CAASPP System 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 the pursuit of enforcing secure practices, ETS and the OTI strive to safeguard the various processes involved in a test development and administration cycle. Those processes are listed below. The practices related to each of the following processes are discussed in detail in Chapter 5, starting on page 29.

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

Staff at LEAs who are central to the processes include LEA CAASPP coordinators, test site coordinators, test examiners, proctors, and observers. The responsibilities for each of the staff members are included in the *LEA CAASPP and Test Site Coordinator Manual* (CDE, 2014a); see page 34 in Chapter 5 for more information.

### Test Directions

A series of instructions compiled in detailed manuals are provided 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.)

***LEA CAASPP and Test Site Coordinator Manual***—Test administration procedures for LEA CAASPP coordinators and test site coordinators (See page 36 in Chapter 5 for more information.)

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

Training in the form of “CAPA Train-the-Trainer” workshops is available in January and is presented in live workshops and a Webcast, which is later archived. An LEA representative who takes the training can then train test site staff to train CAPA examiners and observers. Video segments that model CAPA task administration are made available during the school year; sample materials that support the training are available all year on the [californiatac.org](http://californiatac.org) Web site, at <http://californiatac.org/training/capa/>.

## Universal Tools, Designated Supports, and Accommodations

All public school students participate in the CAASPP System, including students with disabilities and English learners. Students with an IEP and who have significant cognitive disabilities may take the CAPA when they are unable to take a CST or CMA with or without universal tools, designated supports, and/or 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, 2014b).

As noted on the CDE CAPA Participation Criteria Web page, “Since examiners may adapt the CAPA based on students’ instruction mode, accommodations and modifications do not apply to CAPA.” (CDE, 2014c)

## Scores

The CAPA total test raw scores equal the sum of examinees’ scores on the operational tasks. The total raw scores differ in the score range across different CAPA levels.

Raw scores for Level I range from 0 to 40; for the other CAPA levels, the raw-score range is from 0 to 32. Total test 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 categorizing 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 found in Chapter 7, which starts on page 45.

## 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 students 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. Had 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 50 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 58 in Appendix 7.B. 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 51 provides definitions for the demographic groups included in the tables.

## Equating

### Post-Equating

Prior to the 2013 administration, each CAPA form was equated to a reference form using a linking items nonequivalent groups data collection design and methods based on item response theory (IRT) (Hambleton & Swaminathan, 1985). The “base” or “reference” calibrations for the 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 procedure used for post-equating the CAPA involves three steps: calibration, scaling, and linear transformation. Each of those procedures, as described below, is applied to all CAPA tests.

During the 2014 administration, because the intact test forms were used from last year’s CAPA administration, the raw-to-scale-score conversion tables from the 2013 CAPA administration are directly applied to the 2014 administration.

### Calibration

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

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 is controlled by the constant discrimination parameters.

### Scaling

Prior to the 2013 administration, calibrations of the 2013 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 was 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 was 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 was 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 were calculated using the following formula:

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

where,

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

$n_g$  is the number of intervals/groups,

$\theta_j$  is the mean of the ability estimates that fall in interval  $j$ ,

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

$P_n(\theta_j)$  is the probability of a given score for the transformed new form item at ability  $\theta_j$ , and

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

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

### 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:

$$\text{Scale Score} = (35 - \theta_{\text{proficient}}) \times \left( \frac{35 - 30}{\theta_{\text{proficient}} - \theta_{\text{basic}}} \right) + \left( \frac{35 - 30}{\theta_{\text{proficient}} - \theta_{\text{basic}}} \right) \times \theta \tag{2.2}$$

where,

$\theta$  represents the student ability,

$\theta_{\text{proficient}}$  represents the theta cut score for proficient on the spring 2009 base scale,

and

$\theta_{\text{basic}}$  represents the theta cut score for basic on the spring 2009 base scale.

Complete raw-score-to-scale-score conversion tables for the 2014 CAPA are presented in Table 8.D.1 through Table 8.D.14 in Appendix 8.D, starting on page 127. The raw scores and corresponding transformed 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 Score 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

<b>Content Area</b>	<b>CAPA Level</b>	<b>Far Below Basic</b>	<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>
<b>Science</b>	<b>I</b>	15	16 – 29	30 – 34	35 – 38	39 – 60
	<b>III</b>	15 – 21	22 – 29	30 – 34	35 – 39	40 – 60
	<b>IV</b>	15 – 19	20 – 29	30 – 34	35 – 39	40 – 60
	<b>V</b>	15 – 20	21 – 29	30 – 34	35 – 38	39 – 60

## References

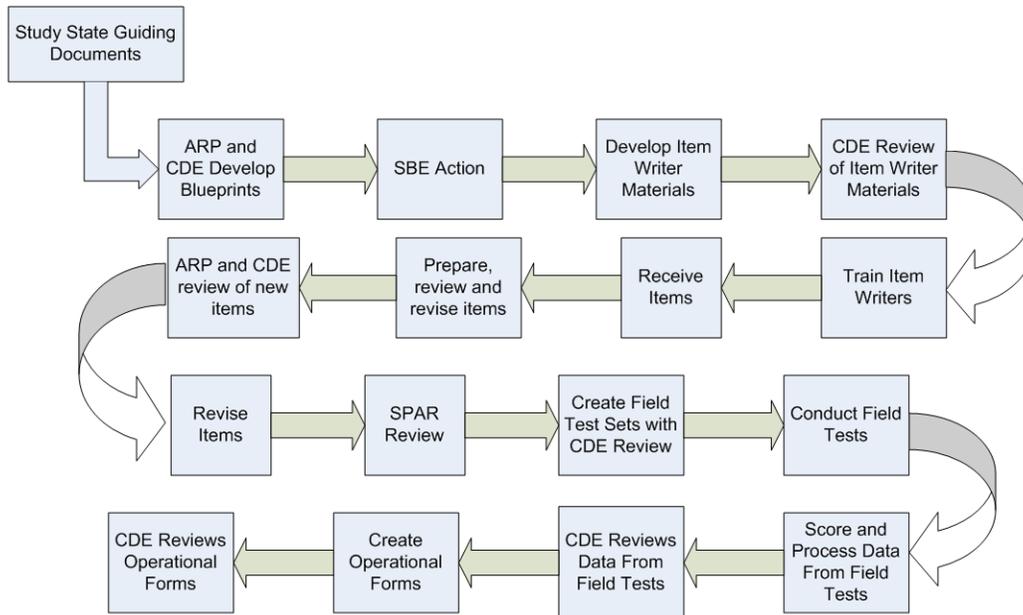
- California Department of Education. (2014a). *2014 LEA CAASPP and test site coordinator manual*. Sacramento, CA. Downloaded from [http://californiatac.org/rsc/pdfs/CAASPP.coord\\_man.2014.pdf](http://californiatac.org/rsc/pdfs/CAASPP.coord_man.2014.pdf)
- California Department of Education. (2014b). *2014 California Alternate Performance Assessment (CAPA) examiner's manual*. Sacramento, CA. Downloaded from [http://californiatac.org/rsc/pdfs/CAPA.examiners\\_manual.nonsecure.2014.pdf](http://californiatac.org/rsc/pdfs/CAPA.examiners_manual.nonsecure.2014.pdf)
- California Department of Education. (2014c). *CAPA participation criteria*. Retrieved from <http://www.cde.ca.gov/TA/tg/sr/participcritria.asp>
- Educational Testing Service. (2002). *ETS standards for quality and fairness*. Princeton, NJ: Author.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston, MA: Kluwer-Nijhoff.
- Linacre, J. M. (2013). *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

Intact test forms from the previous year's test administration were reused during the 2014 administration. This reuse permitted score conversion tables from the previous administration to be used to look up student scores and performance levels. There was no new item (task) development for the 2014 forms.

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

**Figure 3.1 The ETS Item Development Process for the CAASPP System**



### Rules for Task Development

The development of CAPA tasks followed guidelines for task writing approved by the CDE. These guidelines directed 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 were eliminated early in a rigorous task review process when they were only peripherally related to the test and task specifications, did not measure core outcomes reflected in the California content standards, or were not developmentally appropriate.

### Task Specifications

ETS senior content staff led the task writers in the task development and review process. In addition, experienced ETS content specialists and assessment editors reviewed each task during the forms-construction process. The lead assessment specialists for each content area worked 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 followed the SBE-approved item utilization plan to guide the development of the tasks for each content area.

Task specification documents included a description of the constructs to be measured and the California content standards; tasks of the same type should consistently measure the

content standards in the same way each year. The task specifications also provided specific and important guidance to task writers.

The task specifications described the general characteristics of the tasks for each content standard, indicated task types or content to be avoided, and defined the content limits for the tasks. More specifically, the specifications included the following:

- A statement of the strand or topic for the standard
- A full statement of the academic content standard, as found in each 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 contained guidelines for stimulus cards used to assess reading comprehension included 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 developed the item utilization plan for the development of CAPA tasks. The plan included strategies for developing tasks that permitted coverage of all appropriate standards for all tests in each content area and at each grade level. ETS test development staff used this plan to determine the number of tasks to develop for each content area.

The item utilization plan assumed that each year, 25 percent of items on an operational form would be refreshed (replaced); these items would remain in the item bank for future use. The item utilization plan also declared that an additional five percent of the operational items were likely to become unusable because of normal attrition and noted a need to focus development on “critical” standards, those that were difficult to measure well or for which there were few usable items.

For the 2014 CAPA administration, field-test items were repeated as a part of the intact reused form. Detailed information about field testing was presented in the *2013 CAPA Technical Report*.

## **Selection of Task Writers**

### **Criteria for Selecting Task Writers**

The tasks for each CAPA were written by individual task writers with a thorough understanding of the California content standards. Applicants for task writing were screened by senior ETS content staff. Only those with strong content and teaching backgrounds were approved for inclusion in the training program for task writers. Because most of the participants were current or former California educators, they were particularly

knowledgeable about the standards assessed by the CAPA. All task writers met the following minimum qualifications:

- Possession of a bachelor’s degree in the relevant content area or in the field of education with special focus on a particular content of interest; an advanced degree in the relevant content area is desirable
- Previous experience in writing 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 the CAPA underwent an extensive task review process that was 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—currently being reused—at the time the tasks and forms were developed.

### Contractor Review

Once the tasks were written, ETS employed a series of internal reviews. The reviews established the criteria used to judge the quality of the task content and were designed to ensure that each task measured what it was intended to measure. The internal reviews also examined the overall quality of the tasks before they were 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 CAASPP Assessment System, it was essential that many experienced individuals reviewed each task before it was brought to the CDE, the ARPs, and Statewide Pupil Assessment Review (SPAR) panels.

The ETS review process for the CAPA included 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 evaluated 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 were only peripherally related to the test and task specifications, did not measure core outcomes reflected in the California content standards, or were not developmentally appropriate were eliminated early in this rigorous review process.

### 1. Internal Content Review

Test tasks and materials underwent two reviews by the content-area assessment specialists. These assessment specialists made sure that the test tasks and related materials were in compliance with ETS’s written guidelines for clarity, style, accuracy, and appropriateness for California students as well as in compliance with the approved task

specifications. Assessment specialists reviewed 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 was classified with a code for the standard it was intended to measure. The assessment specialists checked all tasks against their classification codes, both to evaluate the correctness of the classification and to ensure that a given task was of a type appropriate to the outcome it was intended to measure. The reviewers could accept the task and classification as written, suggest revisions, or recommend that the task be discarded. These steps occurred prior to the CDE's review.

## **2. Internal Editorial Review**

After the content-area assessment specialists reviewed each task, a group of specially trained editors also reviewed each task in preparation for consideration by the CDE and the ARPs. The editors checked 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 conducted the next level of review. These trained staff members reviewed every task before the CDE and ARP reviews.

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

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

## **Content Expert Reviews**

### **Assessment Review Panels**

ETS was responsible for working with ARPs as tasks were developed for the CAPA. The ARPs are advisory panels to the CDE and ETS and provided guidance on matters related to

task development for the CAPA. The ARPs were responsible for reviewing all newly developed tasks for alignment to the California content standards. The ARPs also reviewed the tasks for accuracy of content, clarity of phrasing, and quality. In their examination of test tasks, the ARPs could raise concerns related to age/level appropriateness and gender, racial, ethnic, and/or socioeconomic bias.

### **Composition of ARPs**

The ARPs comprised current and former teachers, resource specialists, administrators, curricular experts, and other education professionals. Current school staff members met minimum qualifications to serve on the 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 grade 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 ARP committees include representation of genders 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.

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

ARP members were employed as teachers, program specialists, university faculty members, and LEA personnel, had a minimum of a bachelor’s degree, and had experience teaching students, whether in a classroom setting or one-on-one. Due to the use of intact forms in 2014, no field test items were developed. Consequently, no ARP meetings were convened.

### **ARP Meetings for Review of CAPA Tasks**

ETS content-area assessment specialists facilitated the CAPA ARP meetings. Each meeting began with a brief training session on how to review tasks. ETS provided this training, which consisted 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 task specifications
- Overview of criteria for reviewing constructed-response tasks
- Review and evaluation of tasks for bias and sensitivity issues

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

Guidelines for reviewing tasks were 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 respond to 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 reviewed a set of tasks independently and recorded their individual comments. The next step in the review process was for the group to discuss each task. The content-area assessment specialists facilitated the discussion and recorded all recommendations in a master task review booklet. Task review binders and other task evaluation materials also identified potential bias and sensitivity factors for the ARP to consider as a part of its task reviews.

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

### **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 were presented in binders for review. The SPAR panel representatives ensured that the test tasks conformed to the requirements of *EC* Section 60602. If the SPAR panel rejected specific tasks, the tasks were marked for rejection in the item bank and excluded from use on field tests. For the SPAR panel meeting, the task development coordinator was available by telephone to respond to any questions during the course of the meeting.

### **Field Testing**

The primary purposes of field testing are to obtain information about task performance and to obtain statistics that can be used to assemble operational forms. However, because intact 2013 test forms were administered in 2014 and field-test items included in the test forms

were analyzed in the 2013 administration cycle, no additional field test analyses were conducted.

### **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 circumstances, 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 core set of operational tasks and different sets of field-test tasks.

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

The test forms for a given CAPA were distributed by random assignment to LEAs so that a large representative sample of test takers responded to the field-test items embedded in these forms. The random assignment of specific forms ensured that a diverse sample of students took each field-test task. The students did not know which tasks were field-test tasks and which tasks were operational tasks; therefore, their motivation was not expected to vary over the two types of tasks (Patrick & Way, 2008).

## **CDE Data Review**

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

## **Item Banking**

Once the ARP new item (task) review was complete, the tasks were placed in the item bank along with their corresponding review information. Tasks that were accepted by the ARP, SPAR, and CDE were updated to a "field-test ready" status; tasks that were rejected were updated to a "rejected before use" status. ETS then delivered the tasks to the CDE by means of a delivery of the California electronic item bank. Subsequent updates to tasks were 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 did not meet statistical specifications might be rejected; such tasks were updated with a status of "rejected for statistical reasons" and remain unavailable in the bank. These statistics were obtained by the psychometrics group

at ETS, which carefully evaluated each task for its level of difficulty and discrimination as well as conformance to the Rasch partial credit model. Psychometricians also determined if the task functioned similarly for various subgroups of interest.

All unavailable tasks were marked with an availability indicator of “Unavailable,” a reason for rejection as described above, and cause alerts so they are not inadvertently included on subsequent test forms. Status and availability of a task were updated programmatically as tasks were presented for review, accepted or rejected, placed on a form for field-testing, presented for statistical review, and used operationally. All rejection indications were monitored and controlled through ETS’s assessment development processes.

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

## References

- 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 were constructed to measure students' performance relative to California's content standards approved by the SBE. They were also constructed to meet professional standards for validity and reliability. For each CAPA, the content standards and desired psychometric attributes were 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 the California content standards for each level and content area. Also considered is the goal that the tests be short enough so 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. See the *2013 CAPA Technical Report* for more details on the distribution of items at each level and content area.

### Rules for Task Selection

#### Test Blueprints

ETS developed all CAPA 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 CDE STAR CAPA Blueprints Web page at <http://www.cde.ca.gov/ta/tg/sr/capablueprints.asp>.

#### Content Rules and Task Selection

Intact test forms from the 2013 CAPA administration were reused during the 2014 administration. Prior to the 2014 administration, test developers followed a number of rules when developing a new test form for a given grade and content area. First and foremost, they selected tasks that met the blueprint for that grade level and content area. Using an electronic item bank, assessment specialists began by identifying a number of linking tasks. These are tasks that appeared in a previous year's operational administration and were used to equate the administered test forms. Linking tasks were selected to proportionally represent the full blueprint. The selected linking tasks were also reviewed by psychometricians to ensure that the specific psychometric criteria were met.

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

Another consideration was the difficulty of each task. Test developers strived to ensure that the tasks were spread evenly from easy to hard, with some easy and some hard tasks, and a number of tasks in the middle range of difficulty. If tasks did not meet all content and psychometric criteria, staff reviewed the other available tasks to determine if there were other selections that could improve the match of the test to all of the requirements. If such a

match was not attainable, the content team worked in conjunction with psychometricians and the CDE to determine which combination of tasks would best serve the needs of the students taking the test. Chapter 3, starting on page 17, contains further information about this process.

## Psychometric Criteria

The three goals of CAPA test development were as follows:

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

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

Total test assembly targets were developed for each CAPA. These targets were provided to test developers before a test construction cycle began.

### Primary Statistical Targets

The total test targets, or primary statistical targets, used for assembling the CAPA forms for the 2013 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 model, the information curve conditional on each ability level is determined by item difficulty ( $b$ -values) alone. In this case, the test information function (TIF) would, therefore, suffice as the target for conditional test difficulty. Although additional item difficulty targets are not imperative when the target TIF is used for form construction, the target mean and standard deviation of item difficulty ( $b$ -values) consistent with the TIF were still provided to test development staff to help with the test construction process.

The 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 who differ in their ability, and is related to the overall reliability of the test.

### Assembly Targets

The target values for the CAPA, presented in Table 4.1, were used in the 2014 test forms, which are intact test forms developed and used in the 2013 administration. These specifications were developed from the analyses of test forms administered in 2009, the base year in which test results were reported using new scales and new cut scores for the five performance levels: far below basic, below basic, basic, proficient, and advanced.

**Table 4.1 Statistical Targets for CAPA Test Assembly**

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

Content Area	CAPA Level	Target Mean <i>b</i>	Target SD <i>b</i>	Mean AIS	Mean Polyserial
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
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 2014 administration, ETS psychometricians performed a preliminary review of the technical characteristics of the assembled tests. Table 4.2 shows the projected statistical attributes of each CAPA based on the most recent banked item statistics. These values can be compared to the target values in Table 4.1.

Table 4.2 Summary of 2013 CAPA Projected Statistical Attributes

Content Area	CAPA Level	Mean <i>b</i>	SD <i>b</i>	Mean AIS	Min AIS	Max AIS	Mean Polyserial
English–Language Arts	I	-0.59	0.09	3.17	2.87	3.68	0.76
	II	-0.66	0.76	2.30	1.89	3.57	0.73
	III	-0.80	0.42	2.50	2.21	3.27	0.75
	IV	-0.73	0.36	2.24	1.66	2.56	0.77
	V	-0.86	0.47	2.60	2.04	3.12	0.78
Mathematics	I	-0.24	0.14	2.91	2.58	3.30	0.74
	II	-0.99	0.76	2.49	1.24	3.20	0.72
	III	-0.97	0.39	2.49	2.03	3.06	0.70
	IV	-0.65	0.62	2.30	1.50	2.97	0.70
	V	-1.02	0.27	2.57	2.13	2.94	0.74
Science	I	-0.29	0.12	2.90	2.37	3.11	0.78
	III	-1.09	0.42	2.63	2.24	3.04	0.72
	IV	-1.10	0.37	2.69	2.17	3.03	0.68
	V	-0.51	0.62	2.57	1.97	3.42	0.70

### Rules for Task Sequence and Layout

Linking tasks typically were placed in each form first; the sequence of the linking tasks was 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 were 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 CAASPP System are secure documents. For the CAPA administration, every person having access to testing materials maintains the security and confidentiality of the tests. ETS's Code of Ethics requires that all test information, including tangible materials (such as test booklets), confidential files, processes, and activities are kept secure. ETS has systems in place that maintain tight security for test questions and test results as well as for student data. To ensure security for all the tests that ETS develops or handles, ETS maintains an Office of Testing Integrity (OTI), which is described in the next section.

### ETS's Office of Testing Integrity

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

OTI's mission is to:

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

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

### Test Development

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

- Only authorized individuals have access to test content at any step during the 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

As mentioned in Chapter 3, ARP meetings were not held in 2014 because there was no new task development for the 2014 CAPA forms. However, before the 2014 administration,

ETS facilitated ARP meetings every year to review all newly developed CAPA tasks and associated statistics. ETS enforced security measures at ARP meetings to protect the integrity of meeting materials using the following guidelines:

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

### **Item Banking**

When the ARP review was complete, the tasks were placed in the item bank. ETS then delivered the tasks to the CDE through the California electronic item bank. Subsequent updates to content and statistics associated with tasks were based on data collected from field testing and the operational use of the tasks. The latest version of the task is retained in the bank along with the data from every administration that had 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 off site, to prevent loss from a system breakdown or a natural disaster.
- The offsite backup files are kept in secure storage with access limited to authorized personnel only.
- To prevent unauthorized electronic access to the item bank, state-of-the-art network security measures are used.

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

### **Transfer of Forms and 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 server is used as a conduit for the transfer of files; secure test data are not stored permanently on the shared SFTP server.

### **Security of Electronic Files Using a Firewall**

A firewall is software that prevents unauthorized entry to files, e-mail, and other organization-specific programs. 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 Test Management System (CDE, 2014a) remain protected by the ETS firewall software at all times. Due to the sensitive nature of the student

information processed by the Test Management System, the firewall plays a significant role in maintaining an assurance of confidentiality in the users of this information.

## Printing and Publishing

After 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 preapproves all printing vendors before they can work on secured confidential and proprietary testing materials. The printing vendor must submit a completed ETS Printing Plan and a Typesetting Facility Security Plan; both plans document security procedures, access to testing materials, a log of work in progress, personnel procedures, and access to the facilities by the employees and visitors. After reviewing the completed plans, representatives of the OTI visit the printing vendor to conduct an onsite inspection. The printing vendor ships printed test booklets to Pearson and other authorized locations. Pearson distributes the booklets to LEAs in securely packaged boxes.

## Test Administration

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

## Test Delivery

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

- LEA CAASPP coordinators are required to sign and submit a “CAASPP Test Security Agreement for LEA CAASPP Coordinators and CAASPP Test Site Coordinators (For all CAASPP assessments, including field tests)” form to the California Technical Assistance Center before ETS may ship any testing materials to the LEA.
- Test site coordinators have to sign and submit a “CAASPP Test Security Agreement for LEA CAASPP Coordinators and CAASPP Test Site Coordinators (For all CAASPP assessments, including field tests)” form to the LEA CAASPP coordinator before any testing materials may be delivered to the school/test site.
- Anyone having access to the testing materials must sign and submit a “CAASPP Test Security Affidavit for Test Examiners, Proctors, Scribes, and Any Other Persons Having Access to CAASPP Tests (For all CAASPP assessments, including field tests)” form to the test site coordinator before receiving access to any testing materials.
- It is the responsibility of each person participating in the CAASPP System to report immediately any violation or suspected violation of test security or confidentiality. The test site coordinator is responsible for immediately reporting any security violation to the LEA CAASPP coordinator. The LEA CAASPP coordinator must contact the CDE

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

## **Processing and Scoring**

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

- There is controlled access to the facility.
- No 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. School and LEA personnel are not allowed to look at a completed answer document unless required for transcription or to investigate irregular cases.

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

## **Data Management**

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

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

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

## Transfer of Scores via Secure Data Exchange

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

## Statistical Analysis

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

## Reporting and Posting Results

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

- Paper reports, some with individual student results and others with summary results
- A file of individual student results—available for download through the electronic reporting function of the Test Management System’s QTR module—that shows students’ scale scores and performance levels
- Encrypted files of summary results (sent to the CDE by means of SFTP) (Any summary results that have fewer than 11 students are not reported.)
- Task-level statistics based on the results, which are entered into the item bank

## Student Confidentiality

To meet ESEA and state requirements, LEAs must collect demographic data about students. This includes information about students’ ethnicity, parent education, disabilities, whether the student qualifies for the National School Lunch Program (NSLP), and so forth (CDE, 2014b). ETS takes precautions to prevent any of this information from becoming public or being used for anything other than testing purposes. These procedures are applied to all documents in which these student demographic data may appear, including in Pre-ID files and reports.

## Student Test Results

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

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

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

Access to the ETS Computer Processing Center is controlled by employee and visitor identification badges. The Center is secured by doors that can 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 installed in the Center.

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

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

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

## Procedures to Maintain Standardization

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

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

### Test Administrators

The CAPA are administered in conjunction with the other tests that comprise the CAASPP Assessment System. The responsibilities for LEA and test site staff members are included in the *LEA CAASPP and Test Site Coordinator Manual* (CDE, 2014c). This manual is described in the next section.

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

#### LEA CAASPP Coordinator

Each LEA designates an LEA CAASPP coordinator who is responsible for ensuring the proper and consistent administration of the CAASPP tests. LEAs include public school

districts, statewide benefit charter schools, state board–authorized charter schools, county office of education programs, and charter schools testing independently from their home district.

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

### **Test Site Coordinator**

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

Test site coordinators are responsible for making sure that the school has the proper testing materials, distributing testing materials within a school, securing materials before, during, and after the administration period, answering questions from test examiners, preparing and packaging materials to be returned to the LEA after testing, and returning the materials to the LEA. (CDE, 2014c)

### **Test Examiner**

The CAPA are administered to students individually by test examiners who may be assisted by test proctors and scribes. A test examiner is an employee of an LEA or an employee of a nonpublic, nonsectarian school (NPS) who has been trained to administer the tests and has signed a CAASPP Test Security Affidavit. For the CAPA, the test examiner must be a certificated or licensed school staff member (5 CCR Section 850 [w]). Test examiners must follow the directions in the *CAPA Examiner’s Manual* (CDE, 2014d) exactly.

### **Test Proctor**

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

### **Observer**

To establish scoring reliability, the test site coordinator and principal of the school should objectively and randomly select 10 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 [w]) 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, 2014d).

## LEA CAASPP 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 *LEA CAASPP and Test Site Coordinator Manual* contributes to this goal by providing information about the responsibilities of LEA and test site coordinators, as well as those of the other staff involved in the administration cycle (CDE, 2014c). However, the manual is not intended as a substitute for the *CCR, Title 5, Education (5 CCR)*, or to detail all of the coordinator's responsibilities.

## Test Management System Manuals

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

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

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

All public school students participate in the CAASPP Assessment System, 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. The ADA authorizes that test takers with disabilities may be tested under standard conditions if ETS determines that only minor adjustments to the testing environment are required (e.g., wheelchair access, large-print test book, a sign language interpreter for spoken directions).

### Identification

Most students with disabilities and most English learners take the Smarter Balanced for ELA and mathematics and the CST for Science 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 universal tools, designated supports, and accommodations. The "Universal Tools, Designated Supports, and Accommodations for the California Assessment of Student Performance and Progress" for administrations of

California statewide assessments are available on the CDE’s Web site (CDE, 2014g). Because examiners may adapt the CAPA in light of a student’s instructional mode, universal tools, designated supports, and accommodations 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 include 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 Examiner’s Manual* (CDE, 2014d), on page 23 of the nonsecure manual.

## Scoring

CAPA tasks are scored using a 5-point holistic rubric (Level I) or a 4-point holistic rubric (Levels II–V) 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 10 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.

## Testing Incidents

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

The LEA CAASPP coordinator is responsible for immediately notifying the CDE of any irregularities or breaches that occur before, during, or after testing. The test examiner is responsible for immediately notifying the LEA CAASPP coordinator of any security breaches or testing irregularities that occur in the administration of the test. Once the LEA CAASPP coordinator and the CDE have determined that an irregularity or breach has occurred, the CDE instructs the LEA CAASPP coordinator on how and where to identify the irregularity or breach on the student answer document. The information and procedures to assist in identifying incidents and notifying the CDE are provided in the *LEA CAASPP and Test Site Coordinator Manual* (CDE, 2014c).

## Social Media Security Breaches

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

There were no social media security breaches associated with the CAPA in 2014.

## Testing Improprieties

A testing impropriety is any event that occurs before, during, or after test administrations that does not conform to the instructions stated in the *DFAs* (CDE, 2014d) and the *LEA CAASPP and Test Site Coordinator Manual* (CDE, 2014c). These events include test administration errors, disruptions, and student cheating. Testing improprieties generally do not affect test results and are not reported to the CDE or the CAASPP System testing contractor. The CAASPP test site coordinator should immediately notify the LEA CAASPP coordinator of any testing improprieties that occur. It is recommended by the CDE that LEAs and schools maintain records of testing improprieties.

## References

- California Department of Education. (2014a). *2014 Test Management System*. Sacramento, CA. <http://californiatac.org/administration/tms/>
- California Department of Education. (2014b). *2014 CAASPP Pre-ID and Extended Pre-ID Data Corrections instructions manual*. Sacramento, CA. Downloaded from [http://www.startest.org/pdfs/calif-tac.pre-id\\_xdc\\_manual.2014.pdf](http://www.startest.org/pdfs/calif-tac.pre-id_xdc_manual.2014.pdf)
- California Department of Education. (2014c). *2014 CAASPP LEA and test site coordinator manual*. Sacramento, CA. Downloaded from [http://californiatac.org/rsc/pdfs/CAASPP.coord\\_man.2014.pdf](http://californiatac.org/rsc/pdfs/CAASPP.coord_man.2014.pdf)
- California Department of Education (2014d). *2014 California Alternate Performance Assessment (CAPA) examiner's manual*. Sacramento, CA. Downloaded from [http://californiatac.org/rsc/pdfs/CAPA.examiners\\_manual.nonsecure.2014.pdf](http://californiatac.org/rsc/pdfs/CAPA.examiners_manual.nonsecure.2014.pdf)
- California Department of Education. (2014e). *2014 CAASPP Test Administration Setup manual*. Sacramento, CA. Downloaded from [http://www.startest.org/pdfs/STAR.test\\_admin\\_setup.2014.pdf](http://www.startest.org/pdfs/STAR.test_admin_setup.2014.pdf)
- California Department of Education. (2014f). *2014 CAASPP Order Management manual*. Sacramento, CA. Downloaded from [http://www.startest.org/pdfs/calif-tac.order\\_mgmt.2014.pdf](http://www.startest.org/pdfs/calif-tac.order_mgmt.2014.pdf)
- California Department of Education. (2014g). *Universal Tools, Designated Supports, and Accommodations for the California Assessment of Student Performance and Progress*. Sacramento, CA. <http://www.cde.ca.gov/ta/tg/ai/caasppmatrix1.asp>

# Chapter 6: Performance Standards

---

## Background

The CAPA were first administered as a part California's standardized testing program in 2003. Subsequently, the CAPA were revised to better link these tests 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 each performance level (policy-level descriptors) and the associated 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 selected based on the following characteristics:

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

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 would have had

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 these steps, which include training and practice prior to making judgments:

1. Prior to attending the workshop, all panelists received a pre-workshop assignment. The task was to review, on their own, the content standards upon which the 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 that 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 it.
6. After training in the method was complete and confirmed through an evaluation questionnaire, panelists made individual judgments. Working in small groups, they discussed feedback related to other panelists’ judgments and feedback based on student performance data (impact data). Note that no impact data were presented to the Level I panel due to the change in the Level I rubric. Panelists could revise their judgments during the process if they wished.
7. The final recommended cut scores were based on an average of panelists’ judgment scores 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 adoption.

### **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 lists 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 were 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 was 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 test level and content area. They were 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 was mapped to an IRT *theta* ( $\theta$ ) using the test characteristic function and then transformed to the scale score metric using the following equation:

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

where,

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

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

$\theta_{\text{basic}}$  represents the theta corresponding to the cut score for basic.

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 50 in Chapter 7 presents the percentages of examinees meeting each performance level in 2014.

## 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. These standards dictate that ETS provides accurate and understandable assessment results to the intended recipients. It is also ETS's mission to provide appropriate guidelines for score interpretation and cautions about the limitations in the meaning and use of the test scores. Finally, ETS conducts analyses needed to ensure that the assessments are equitable for various groups of test-takers.

### Procedures for Maintaining and Retrieving Individual Scores

Each CAPA is composed entirely of performance tasks; every 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 CAASPP summary reports and defines the appropriate codes to use when a student does not take or complete a test or when a score will not be reported
- **Score Key and Score 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 LEAs to Pearson and defines the action(s) to be taken in response
- **Edit Specifications**—Describes edits, defaults, and solutions to errors encountered while data are being captured as answer documents are processed including matching observer documents to examiner documents

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, including those available via electronic reporting, must comply with the ETS scoring specifications. Pearson has quality control checks in place to ensure the quality and accuracy of scanning and the transfer of scores into the database of student records.

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

## Types of Scores

### Raw Score

For the CAPA for ELA and mathematics, there are five test levels and eight operational tasks per level. For the CAPA for science, there are four test levels and eight operational 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. For all CAPA tests, the total test raw score equals the sum of the eight operational task scores. The raw scores for Level I range from 0 to 40; for the other CAPA levels, the raw score 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 13. 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, 2014a).

## 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

After Pearson applied 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 150.

## 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 describe 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
<b>Scale Score Information</b>					
Number of examinees	14,922	5,872	6,968	10,134	10,368
Mean score	41.45	38.29	39.44	38.95	38.66
SD *	10.85	6.24	5.98	8.18	6.56
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	41.00	38.00	40.00	40.00	39.00
Reliability	0.88	0.84	0.89	0.89	0.90
SEM †	3.68	2.47	2.01	2.74	2.08
<b>Raw Score Information</b>					
Mean score	25.84	18.13	19.88	18.19	20.48
SD *	11.55	6.04	6.78	7.27	6.99
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	28.00	18.00	21.00	19.00	21.00
Reliability	0.88	0.84	0.89	0.89	0.90
SEM †	3.92	2.39	2.27	2.43	2.22

Level	I	II	III	IV	V
<b>Task Information</b>					
Number of tasks	8	8	8	8	8
Mean AIS ‡	3.24	2.27	2.49	2.28	2.57
SD AIS ‡	0.26	0.56	0.34	0.27	0.39
Min. AIS	2.85	1.87	2.23	1.68	2.09
Max. AIS	3.60	3.56	3.28	2.50	3.09
Possible range	0-5	0-4	0-4	0-4	0-4
Mean polyserial	0.79	0.74	0.79	0.79	0.81
SD polyserial	0.05	0.08	0.09	0.08	0.06
Min. polyserial	0.67	0.62	0.66	0.61	0.69
Max. polyserial	0.84	0.85	0.87	0.84	0.86

\* 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
<b>Scale Score Information</b>					
Number of examinees	14,866	5,864	6,952	10,103	10,324
Mean score	36.61	37.25	36.26	36.55	37.27
SD *	9.52	8.66	5.86	7.67	7.95
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	37.00	37.00	36.00	37.00	38.00
Reliability	0.86	0.85	0.84	0.84	0.87
SEM †	3.50	3.34	2.37	3.05	2.86
<b>Raw Score Information</b>					
Mean score	23.74	20.01	19.78	18.26	20.12
SD *	11.16	6.36	6.38	6.60	7.34
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	25.00	20.00	20.00	19.00	21.00
Reliability	0.86	0.85	0.84	0.84	0.87
SEM †	4.11	2.45	2.57	2.63	2.64
<b>Task Information</b>					
Number of tasks	8	8	8	8	8
Mean AIS ‡	2.98	2.51	2.48	2.30	2.53
SD AIS ‡	0.31	0.62	0.38	0.60	0.32
Min. AIS	2.64	1.25	2.02	1.46	2.09
Max. AIS	3.34	3.21	3.11	2.89	2.88
Possible range	0-5	0-4	0-4	0-4	0-4
Mean polyserial	0.77	0.76	0.73	0.75	0.78
SD polyserial	0.04	0.07	0.11	0.09	0.04
Min. polyserial	0.71	0.65	0.55	0.64	0.71
Max. polyserial	0.82	0.86	0.84	0.87	0.84

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

**Table 7.4 Summary Statistics Describing Student Scores: Science**

Level	I	III	IV	V
<b>Scale Score Information</b>				
Number of examinees	3,800	3,551	3,290	3,450
Mean score	37.61	36.09	35.73	35.86
SD *	11.14	4.65	5.69	5.12
Possible range	15–60	15–60	15–60	15–60
Obtained range	15–60	15–60	15–60	15–60
Median	37.00	36.00	36.00	36.00
Reliability	0.90	0.84	0.86	0.85
SEM †	3.58	1.85	2.10	2.01
<b>Raw Score Information</b>				
Mean score	24.47	20.91	21.32	20.18
SD *	11.93	5.69	6.40	5.96
Possible range	0–40	0–32	0–32	0–32
Obtained range	0–40	0–32	0–32	0–32
Median	26.00	21.00	22.00	21.00
Reliability	0.90	0.84	0.86	0.85
SEM †	3.83	2.27	2.35	2.33
<b>Task Information</b>				
Number of tasks	8	8	8	8
Mean AIS ‡	3.07	2.63	2.68	2.53
SD AIS ‡	0.21	0.31	0.27	0.57
Min. AIS	2.57	2.21	2.20	1.94
Max. AIS	3.28	3.17	2.94	3.36
Possible range	0–5	0–4	0–4	0–4
Mean polyserial	0.81	0.73	0.75	0.76
SD polyserial	0.03	0.04	0.04	0.05
Min. polyserial	0.75	0.65	0.70	0.66
Max. polyserial	0.84	0.79	0.81	0.82

\* 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 P1 data file was used for the analyses in this chapter. This file contained data collected from all LEAs but did not include corrections of demographic data through CALPADS. In addition, students with invalid scores were excluded from the tabled results.

**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%	7%	25%	58%
	II	1%	5%	15%	37%	42%
	III	2%	2%	11%	34%	50%
	IV	3%	7%	16%	33%	41%
	V	2%	3%	16%	35%	44%
Mathematics	I	7%	9%	16%	31%	38%
	II	3%	14%	18%	32%	33%

Content Area	CAPA Level	Far Below Basic	Below Basic	Basic	Proficient	Advanced
Mathematics (continued)	III	1%	7%	29%	37%	27%
	IV	2%	11%	22%	35%	30%
	V	2%	9%	17%	34%	37%
Science	I	8%	8%	17%	25%	41%
	III	1%	4%	25%	54%	16%
	IV	1%	9%	25%	48%	18%
	V	1%	5%	30%	41%	24%

Table 7.A.1 through Table 7.A.3 in Appendix 7.A, starting on page 56, 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 starting on page 58 in Table 7.B.1 through Table 7.B.3 for the CAPA.

In these 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. Percentages in these tables may not sum up to 100 due to rounding.

**Table 7.6 Subgroup Definitions**

Subgroup	Definition
Gender	<ul style="list-style-type: none"> <li>• Male</li> <li>• Female</li> </ul>
Ethnicity	<ul style="list-style-type: none"> <li>• African American</li> <li>• American Indian or Alaska Native</li> <li>• Asian                             <ul style="list-style-type: none"> <li>– Asian Indian</li> <li>– Cambodian</li> <li>– Chinese</li> <li>– Hmong</li> <li>– Japanese</li> <li>– Korean</li> <li>– Laotian</li> <li>– Vietnamese</li> <li>– Other Asian</li> </ul> </li> <li>• Hispanic or Latino</li> <li>• Pacific Islander                             <ul style="list-style-type: none"> <li>– Guamanian</li> </ul> </li> </ul>

Subgroup	Definition
	<ul style="list-style-type: none"> <li>– Native Hawaiian</li> <li>– Samoan</li> <li>– Tahitian</li> <li>– Other Pacific Islander</li> <li>• Filipino</li> <li>• White (not Hispanic)</li> </ul>
English-language Fluency	<ul style="list-style-type: none"> <li>• English only</li> <li>• Initially fluent English proficient</li> <li>• English learner</li> <li>• Reclassified fluent English proficient</li> </ul>
Economic Status	<ul style="list-style-type: none"> <li>• Not economically disadvantaged</li> <li>• Economically disadvantaged</li> </ul>
Primary Disability	<ul style="list-style-type: none"> <li>• Mental retardation/Intellectual disability</li> <li>• Hard of hearing</li> <li>• Deafness</li> <li>• Speech or language impairment</li> <li>• Visual impairment</li> <li>• Emotional disturbance</li> <li>• Orthopedic impairment</li> <li>• Other health impairment</li> <li>• Specific learning impairment</li> <li>• Deaf-blindness</li> <li>• Multiple disabilities</li> <li>• Autism</li> <li>• Traumatic brain injury</li> </ul>

## Reports Produced and Scores for Each Report

The tests that make up the CAASPP Assessment System provide results or score summaries that are reported for different purposes. The three major purposes are:

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

Providing data for federal accountability programs for schools and LEAS—AYP data are submitted to the USDOE for the CAPA for ELA and mathematics in grade ten. A detailed description of the uses and applications of CAASPP reports is presented in the next section.

### Types of Score Reports

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

**Table 7.7 Types of CAPA Reports**

1. Summary Reports	<ul style="list-style-type: none"> <li>▪ CAASPP Student Master List Summary</li> <li>▪ CAASPP Subgroup Summary (including Ethnicity for Economic Status)</li> </ul>
2. Individual Reports	<ul style="list-style-type: none"> <li>▪ CAASPP Student Record Label</li> <li>▪ CAASPP Student Master List</li> <li>▪ CAASPP Student Report for the CAPA</li> </ul>

3. Internet Reports	<ul style="list-style-type: none"> <li>▪ CAPA Scores (state, county, district, school)</li> <li>▪ CAPA Summary Scores (state, county, LEA, school)</li> </ul>
---------------------	---

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

### Score Report Contents

The CAASPP 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 CAASPP Student Report and the other reports is provided in Appendix 7.C on page 64.

### Score Report Applications

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

Schools may use the 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 LEAs and schools identify strengths and weaknesses in their instructional programs. Each year, LEAs 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 for ELA and mathematics in grade ten are used for federal accountability programs to monitor each school's and LEA's progress toward achieving established goals: CAPA results 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 AYP report. Each year, California schools and LEAs must meet AYP goals by showing that a specified percentage of CAPA test-takers at the district and school levels are performing at or above the proficient level on the CAPA for ELA and mathematics.

## Criteria for Interpreting Test Scores

An LEA 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, 2014b). 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 *2014 CAASPP Post-Test Guide* (CDE, 2014c).

## References

- California Department of Education. (2014a). *2014 CAPA examiner's manual*. Sacramento, CA. Downloaded from [http://californiatac.org/rsc/pdfs/CAPA.examiners\\_manual.nonsecure.2014.pdf](http://californiatac.org/rsc/pdfs/CAPA.examiners_manual.nonsecure.2014.pdf)
- California Department of Education. (2014b). *2014 CAASPP CST/CMA, CAPA, and STS printed reports*. Sacramento, CA. Downloaded from <http://californiatac.org/rsc/pdfs/CAASPP.reports.2014.pdf>
- California Department of Education. (2014c). *2014 CAASPP post-test guide*. Sacramento, CA. Downloaded from [http://californiatac.org/rsc/pdfs/CAASPP.post-test\\_guide.2014.pdf](http://californiatac.org/rsc/pdfs/CAASPP.post-test_guide.2014.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

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

**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	1,834	12.29	27	0.46	76	1.09	113	1.12	213	2.05
57–59	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
54–56	663	4.44	32	0.54	N/A	N/A	148	1.46	N/A	N/A
51–53	N/A	N/A	47	0.80	135	1.94	237	2.34	234	2.26
48–50	700	4.69	213	3.63	199	2.86	613	6.05	N/A	N/A
45–47	1,845	12.36	334	5.69	509	7.30	1,204	11.88	776	7.48
42–44	2,025	13.57	1,108	18.87	1,469	21.08	1,849	18.25	1,661	16.02
39–41	2,473	16.57	1,103	18.78	1,820	26.12	1,770	17.47	2,815	27.15
36–38	2,276	15.25	1,462	24.90	1,173	16.83	1,233	12.17	1,721	16.60
33–35	973	6.52	819	13.95	935	13.42	1,157	11.42	1,412	13.62
30–32	543	3.64	370	6.30	375	5.38	830	8.19	1,001	9.65
27–29	450	3.02	131	2.23	96	1.38	429	4.23	194	1.87
24–26	158	1.06	50	0.85	68	0.98	97	0.96	94	0.91
21–23	N/A	N/A	62	1.06	58	0.83	85	0.84	96	0.93
18–20	166	1.11	40	0.68	19	0.27	81	0.80	48	0.46
15–17	816	5.47	74	1.26	36	0.52	288	2.84	103	0.99

**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	840	5.65	94	1.60	46	0.66	70	0.69	350	3.39
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	96	1.64	N/A	N/A	83	0.82	N/A	N/A
51–53	N/A	N/A	194	3.31	84	1.21	134	1.33	N/A	N/A
48–50	365	2.46	355	6.05	N/A	N/A	206	2.04	362	3.51
45–47	299	2.01	337	5.75	165	2.37	572	5.66	343	3.32
42–44	1,565	10.53	596	10.16	484	6.96	1,365	13.51	1,316	12.75
39–41	2,613	17.58	953	16.25	1,536	22.09	1,773	17.55	1,994	19.31
36–38	3,374	22.70	931	15.88	1,764	25.37	1,900	18.81	2,193	21.24
33–35	2,462	16.56	589	10.04	1,321	19.00	1,158	11.46	1,566	15.17
30–32	1,072	7.21	746	12.72	1,020	14.67	1,460	14.45	985	9.54
27–29	806	5.42	405	6.91	243	3.50	410	4.06	323	3.13
24–26	184	1.24	252	4.30	95	1.37	310	3.07	297	2.88
21–23	148	1.00	80	1.36	72	1.04	294	2.91	245	2.37
18–20	N/A	N/A	78	1.33	28	0.40	77	0.76	73	0.71
15–17	1,138	7.66	158	2.69	94	1.35	291	2.88	277	2.68

**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	414	10.89	32	0.90	50	1.52	46	1.33
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	117	3.08	N/A	N/A	N/A	N/A	N/A	N/A
45–47	87	2.29	45	1.27	75	2.28	50	1.45
42–44	439	11.55	184	5.18	110	3.34	133	3.86
39–41	514	13.53	498	14.02	551	16.75	589	17.07
36–38	767	20.18	1,248	35.15	960	29.18	1,163	33.71
33–35	568	14.95	980	27.60	853	25.93	929	26.93
30–32	276	7.26	393	11.07	367	11.16	327	9.48
27–29	86	2.26	88	2.48	196	5.96	123	3.57
24–26	153	4.03	42	1.18	52	1.58	27	0.78
21–23	35	0.92	21	0.59	17	0.52	12	0.35
18–20	45	1.18	4	0.11	20	0.61	19	0.55
15–17	299	7.87	16	0.45	39	1.19	32	0.93

## Appendix 7.B—Demographic Summaries

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

	Number Tested	Percentage in Performance Level				
		Far Below Basic	Below Basic	Basic	Proficient	Advanced
All valid scores	48,264	3%	5%	12%	32%	48%
Male	31,768	3%	5%	12%	32%	48%
Female	16,201	3%	5%	12%	31%	48%
Gender unknown	295	4%	6%	14%	31%	45%
American Indian	389	2%	4%	9%	31%	55%
Asian American	3,364	4%	6%	14%	35%	41%
Pacific Islander	242	3%	4%	15%	35%	43%
Filipino	1,469	3%	6%	15%	34%	42%
Hispanic	25,643	3%	5%	12%	31%	49%
African American	4,320	3%	4%	12%	31%	51%
White	10,748	3%	5%	12%	31%	49%
Ethnicity unknown	2,089	3%	6%	14%	32%	45%
English only	28,292	3%	5%	13%	32%	48%
Initially fluent English proficient	727	5%	8%	15%	32%	41%
English learner	15,809	3%	5%	12%	32%	49%
Reclassified fluent English proficient	2,103	2%	4%	12%	31%	51%
English proficiency unknown	1,333	3%	5%	10%	30%	51%
Mental retardation/Intellectual disability	18,438	2%	4%	13%	33%	47%
Hard of hearing	303	2%	3%	12%	33%	50%
Deafness	291	2%	6%	18%	40%	34%
Speech or language impairment	1,342	0%	1%	4%	26%	69%
Visual impairment	387	7%	10%	10%	29%	44%
Emotional disturbance	313	1%	1%	4%	20%	75%
Orthopedic impairment	3,050	6%	8%	12%	32%	41%
Other health impairment	2,022	2%	3%	7%	28%	60%
Specific learning impairment	2,656	0%	0%	2%	17%	81%
Deaf-blindness	38	8%	16%	16%	24%	37%
Multiple disabilities	2,835	8%	11%	14%	31%	36%
Autism	15,144	3%	5%	14%	33%	44%
Traumatic brain injury	309	5%	6%	8%	31%	50%
Unknown	1,136	3%	5%	10%	33%	49%
Not economically disadvantaged	15,653	4%	6%	14%	33%	44%
Economically disadvantaged	29,955	2%	4%	12%	31%	51%
Economic status unknown	2,656	3%	4%	10%	32%	51%
<b>Primary Ethnicity—Not Economically Disadvantaged</b>						
American Indian	122	3%	7%	11%	39%	39%
Asian American	1,872	4%	6%	15%	35%	40%
Pacific Islander	88	7%	7%	16%	32%	39%
Filipino	892	3%	7%	16%	33%	41%
Hispanic	4,748	4%	7%	13%	32%	44%
African American	1,196	3%	5%	16%	33%	42%
White	6,017	3%	5%	14%	32%	45%
Ethnicity unknown	718	3%	6%	14%	32%	45%

	Number Tested	Percentage in Performance Level				
		Far Below Basic	Below Basic	Basic	Proficient	Advanced
<b>Primary Ethnicity—Economically Disadvantaged</b>						
American Indian	249	1%	2%	8%	27%	62%
Asian American	1,310	4%	6%	15%	33%	42%
Pacific Islander	136	1%	2%	15%	38%	45%
Filipino	519	3%	6%	15%	33%	43%
Hispanic	19,823	2%	4%	12%	31%	50%
African American	2,853	2%	4%	10%	30%	53%
White	4,147	2%	4%	10%	29%	54%
Ethnicity unknown	918	2%	6%	15%	33%	44%
<b>Primary Ethnicity—Unknown Economic Status</b>						
American Indian	18	0%	6%	0%	28%	67%
Asian American	182	4%	5%	10%	40%	41%
Pacific Islander	18	0%	6%	11%	28%	56%
Filipino	58	3%	3%	5%	50%	38%
Hispanic	1,072	3%	5%	9%	32%	51%
African American	271	3%	2%	13%	24%	58%
White	584	3%	4%	10%	33%	50%
Ethnicity unknown	453	3%	4%	13%	30%	50%

\* Results for groups with 10 or fewer members are not reported.

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

	Number Tested	Percentage in Performance Level				
		Far Below Basic	Below Basic	Basic	Proficient	Advanced
All valid scores	48,109	3%	10%	19%	33%	34%
Male	31,671	3%	9%	19%	33%	36%
Female	16,145	4%	11%	21%	33%	31%
Gender unknown	293	4%	10%	20%	33%	33%
American Indian	386	2%	7%	18%	32%	40%
Asian American	3,354	4%	10%	20%	33%	32%
Pacific Islander	239	3%	13%	20%	31%	33%
Filipino	1,467	3%	11%	19%	33%	33%
Hispanic	25,565	3%	10%	19%	33%	35%
African American	4,295	3%	9%	19%	35%	34%
White	10,719	4%	10%	21%	33%	33%
Ethnicity unknown	2,084	4%	11%	20%	33%	32%
English only	28,198	4%	10%	20%	33%	33%
Initially fluent English proficient	726	5%	11%	21%	33%	30%
English learner	15,763	3%	10%	18%	33%	35%
Reclassified fluent English proficient	2,098	2%	8%	18%	32%	40%
English proficiency unknown	1,324	4%	8%	18%	33%	37%
Mental retardation/Intellectual disability	18,390	2%	11%	23%	34%	30%
Hard of hearing	302	2%	7%	17%	36%	37%
Deafness	289	2%	4%	9%	38%	47%
Speech or language impairment	1,340	0%	3%	10%	37%	50%
Visual impairment	384	13%	13%	21%	27%	26%
Emotional disturbance	308	2%	2%	6%	32%	58%
Orthopedic impairment	3,033	9%	15%	22%	30%	24%
Other health impairment	2,021	3%	7%	17%	32%	41%
Specific learning impairment	2,651	0%	1%	5%	30%	64%
Deaf-blindness	36	11%	31%	8%	39%	11%
Multiple disabilities	2,821	11%	16%	23%	28%	22%
Autism	15,098	3%	9%	19%	35%	35%
Traumatic brain injury	309	6%	10%	20%	31%	33%
Unknown	1,127	4%	10%	18%	32%	36%
Not economically disadvantaged	15,610	4%	11%	21%	33%	30%
Economically disadvantaged	29,867	3%	9%	19%	33%	36%
Economic status unknown	2,632	3%	8%	17%	34%	37%
<b>Primary Ethnicity—Not Economically Disadvantaged</b>						
American Indian	122	4%	8%	25%	34%	28%
Asian American	1,867	4%	11%	20%	33%	32%
Pacific Islander	88	5%	16%	16%	26%	38%
Filipino	891	4%	12%	21%	32%	31%
Hispanic	4,736	5%	12%	21%	33%	28%
African American	1,189	5%	11%	20%	34%	30%
White	6,002	4%	11%	22%	33%	30%
Ethnicity unknown	715	4%	11%	21%	35%	29%

	Number Tested	Percentage in Performance Level				
		Far Below Basic	Below Basic	Basic	Proficient	Advanced
<b>Primary Ethnicity—Economically Disadvantaged</b>						
American Indian	247	2%	7%	14%	31%	46%
Asian American	1,307	5%	10%	20%	32%	33%
Pacific Islander	133	3%	11%	23%	34%	30%
Filipino	518	3%	11%	18%	34%	34%
Hispanic	19,762	3%	9%	18%	33%	36%
African American	2,841	3%	8%	19%	36%	35%
White	4,141	3%	8%	19%	33%	38%
Ethnicity unknown	918	3%	12%	22%	29%	34%
<b>Primary Ethnicity—Unknown Economic Status</b>						
American Indian	17	0%	0%	18%	35%	47%
Asian American	180	4%	9%	21%	36%	30%
Pacific Islander	18	0%	11%	17%	33%	39%
Filipino	58	3%	5%	10%	40%	41%
Hispanic	1,067	3%	8%	17%	34%	38%
African American	265	3%	9%	15%	32%	41%
White	576	3%	9%	19%	33%	36%
Ethnicity unknown	451	3%	8%	17%	37%	35%

\* Results for groups with 10 or fewer members are not reported.

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

	Number Tested	Percentage in Performance Level				
		Far Below Basic	Below Basic	Basic	Proficient	Advanced
All valid scores	14,091	3%	6%	24%	42%	25%
Male	9,195	3%	6%	24%	41%	26%
Female	4,846	3%	6%	24%	42%	24%
Gender Unknown	50	8%	6%	24%	32%	30%
American Indian	119	2%	5%	16%	45%	32%
Asian American	961	3%	9%	28%	39%	21%
Pacific Islander	68	1%	9%	22%	46%	22%
Filipino	452	3%	9%	25%	41%	21%
Hispanic	7,408	3%	6%	23%	42%	26%
African American	1,293	3%	6%	24%	44%	23%
White	3,237	3%	6%	23%	41%	26%
Ethnicity unknown	553	2%	8%	28%	38%	24%
English only	8,297	3%	7%	24%	42%	25%
Initially fluent English proficient	236	5%	7%	29%	38%	21%
English learner	4,506	3%	6%	23%	42%	25%
Reclassified fluent English proficient	771	2%	5%	26%	41%	27%
English proficiency unknown	281	4%	7%	26%	34%	29%
Mental retardation/Intellectual disability	5,635	2%	6%	26%	43%	24%
Hard of hearing	97	3%	5%	19%	52%	22%
Deafness	91	1%	5%	16%	60%	16%
Speech or language impairment	321	0%	1%	13%	57%	30%
Visual impairment	147	16%	5%	26%	33%	20%
Emotional disturbance	99	1%	3%	10%	35%	51%
Orthopedic impairment	900	9%	11%	24%	35%	20%
Other health impairment	572	3%	4%	18%	46%	30%
Specific learning impairment	881	0%	0%	5%	46%	49%
Deaf-blindness	10	—	—	—	—	—
Multiple disabilities	851	11%	12%	26%	31%	20%
Autism	4,151	2%	8%	27%	41%	22%
Traumatic brain injury	91	4%	7%	18%	46%	25%
Unknown	245	4%	6%	27%	38%	25%
Not economically disadvantaged	4,710	4%	8%	27%	40%	21%
Economically disadvantaged	8,844	3%	5%	22%	43%	27%
Economic status unknown	537	4%	6%	25%	36%	29%
<b>Primary Ethnicity—Not Economically Disadvantaged</b>						
American Indian	46	4%	11%	17%	50%	17%
Asian American	551	3%	11%	29%	35%	21%
Pacific Islander	18	6%	17%	22%	39%	17%
Filipino	292	2%	9%	25%	41%	22%
Hispanic	1,357	4%	8%	28%	40%	20%
African American	371	6%	8%	26%	43%	17%
White	1,864	3%	8%	26%	40%	23%
Ethnicity unknown	211	3%	6%	29%	39%	22%

	Percentage in Performance Level					
	Number Tested	Far Below Basic	Below Basic	Basic	Proficient	Advanced
<b>Primary Ethnicity—Economically Disadvantaged</b>						
American Indian	71	0%	1%	15%	42%	41%
Asian American	383	3%	7%	26%	44%	20%
Pacific Islander	46	0%	7%	20%	48%	26%
Filipino	149	4%	10%	27%	40%	19%
Hispanic	5,839	3%	5%	22%	42%	27%
African American	868	2%	6%	23%	43%	25%
White	1,240	2%	4%	20%	44%	31%
Ethnicity unknown	248	1%	9%	26%	40%	24%
<b>Primary Ethnicity—Unknown Economic Status</b>						
American Indian	2	—	—	—	—	—
Asian American	27	7%	4%	30%	30%	30%
Pacific Islander	4	—	—	—	—	—
Filipino	11	18%	0%	9%	36%	36%
Hispanic	212	3%	6%	26%	37%	28%
African American	54	2%	0%	20%	46%	31%
White	133	4%	7%	24%	35%	30%
Ethnicity unknown	94	4%	11%	28%	29%	29%

\* Results for groups with 10 or fewer members are not reported.

## Appendix 7.C—Types of Score Reports

Table 7.C.1 Score Reports Reflecting CAPA Results

2014 CAASPP CAPA Student Reports	
Description	Distribution
<b>The CAPA Student Report</b>	
<p>This report provides parents/guardians and teachers with the student's results, presented in tables and graphs.</p> <p>Data presented include the following:</p> <ul style="list-style-type: none"> <li>• Scale scores</li> <li>• Performance levels (advanced, proficient, basic, below basic, and far below basic)</li> </ul>	<p>This report includes individual student results and is not distributed beyond parents/guardians and the student's school.</p> <p>Two copies of this report are provided for each student. One is for the student's current teacher and one is distributed by the LEA to parents/guardians.</p>
<b>Student Record Label</b>	
<p>These reports are printed on adhesive labels to be affixed to the student's permanent school records. Each student shall have an individual record of accomplishment that includes CAASPP testing results (see California <i>EC</i> Section 60607[a]).</p> <p>Data presented include the following for each content area tested:</p> <ul style="list-style-type: none"> <li>• Scale scores</li> <li>• Performance levels</li> </ul>	<p>This report includes individual student results and is not distributed beyond the student's school.</p>
<b>Student Master List</b>	
<p>This report is an alphabetical roster that presents individual student results. It includes the following data for each CAPA content area tested:</p> <ul style="list-style-type: none"> <li>• Scale scores</li> <li>• Performance levels</li> </ul>	<p>This report provides administrators and teachers with all students' results within each grade or within each grade and year-round schedule at a school.</p> <p>Because this report includes individual student results, it is not distributed beyond the student's school. It is recommended that summary reports be retained until the grade level exits the school.</p>
<b>Student Master List Summary</b>	
<p>This report summarizes student results at the school, district, county, and state levels for each grade. It does not include any individual student information.</p> <p>For each CAPA grade and level, the following data are summarized by content area tested:</p> <ul style="list-style-type: none"> <li>• Number of students enrolled</li> <li>• Number and percent of students tested</li> <li>• Number and percent of valid scores</li> <li>• Number tested with scores</li> <li>• Mean scale score</li> <li>• Scale score standard deviation</li> <li>• Number and percent of students scoring at each performance level</li> </ul>	<p>This report is a resource for evaluators, researchers, teachers, parents/guardians, community members, and administrators.</p> <p>One copy is packaged for the school and one for the LEA.</p> <p>This report is also produced for school districts, counties, and the state.</p> <p><b>Note:</b> The data in this report may be shared with parents/guardians, community members, and the media only if the data are for 11 or more students. It is recommended that summary reports be retained for at least five years.</p>

2014 CAASPP CAPA Student Reports	
Description	Distribution
<b>Subgroup Summary</b>	
<p>This set of reports disaggregates and reports results by the following subgroups:</p> <ul style="list-style-type: none"> <li>• All students</li> <li>• Disability status (Disabilities among CAPA students include specific disabilities.)</li> <li>• Economic status</li> <li>• Gender</li> <li>• English proficiency</li> <li>• Primary ethnicity</li> </ul> <p>These reports contain no individual student-identifying information and are aggregated at the school, district, county, and state levels. CAPA statistics are listed by CAPA level.</p> <p>For each subgroup within a report and for the total number of students, the following data are included for each test:</p> <ul style="list-style-type: none"> <li>• Total number tested in the subgroup</li> <li>• Percent of enrollment tested in the subgroup</li> <li>• Number and percent of valid scores</li> <li>• Number tested who received scores</li> <li>• Mean scale score</li> <li>• Standard deviation of scale score</li> <li>• Number and percent of students scoring at each performance level</li> </ul>	<p>This report is a resource for evaluators, researchers, teachers, parents/guardians, community members, and administrators.</p> <p>One copy is packaged for the school and one for the LEA.</p> <p>This report is also produced for school districts, counties, and the state.</p> <p><b>Note:</b> The data on this report may be shared with parents/guardians, community members, and the media only if the data are for 11 or more students. It is recommended that summary reports be retained for at least five years.</p>

<b>2014 CAASPP CAPA Student Reports</b>	
<b>Description</b>	<b>Distribution</b>
<b>Subgroup Summary—Ethnicity for Economic Status</b>	
<p>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).</p> <p>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.</p> <p>For each subgroup within a report, and for the total number of students, the following data are included:</p> <ul style="list-style-type: none"> <li>• Total number tested in the subgroup</li> <li>• Percent of enrollment tested in the subgroup</li> <li>• Number and percent of valid scores</li> <li>• Number tested who received scores</li> <li>• Mean scale score</li> <li>• Standard deviation of scale score</li> <li>• Number and percent of students scoring at each performance level</li> </ul>	<p>This report is a resource for evaluators, researchers, teachers, parents/guardians, community members, and administrators. One copy is packaged for the school and one for the LEA.</p> <p>This report is also produced for school districts, counties, and the state.</p> <p><b>Note:</b> The data on this report may be shared with parents/guardians, community members, and the media only if the data are for 11 or more students. It is recommended that summary reports be retained for at least five years.</p>

## Chapter 8: Analyses

---

This chapter summarizes the task (item)- and test-level statistics obtained for the CAPA administered during the spring 2014 test administration.

The statistics presented in this chapter are divided into four 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

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

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

1. Appendix 8.A on page 86 presents the classical item analyses, including average item score (AIS) and polyserial correlation coefficient, and associated flags, for the operational 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 69.
2. Appendix 8.B on page 96 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 110 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 127 presents the scoring tables obtained as a result of the IRT equating process after the 2013 administration.
5. Appendix 8.E on page 141 shows the distribution of primary disabilities for students who took each CAPA level and content area.

### Samples Used for the Analyses

CAPA analyses were conducted at different times after test administration and involved varying proportions of the full CAPA data.

During the 2014 administration, neither IRT calibrations nor scaling are implemented because of the intact forms used. The summary statistics describing the samples for 2013 and 2014 are presented in Table 8.1.

For the intact forms without any replacement or edited items, the IRT results for calibration and scaling based on the equating sample of the previous administration can be found in Appendix D of the *2013 CAPA Technical Report*, which is the report for the year each CAPA form was administered originally.

**Table 8.1 CAPA Raw Score Means and Standard Deviations: Total Population for 2013 and 2014**

Content Area	Level	2013			2014		
		N	Mean	SD	N	Mean	SD
English–Language Arts	I	14,707	26.07	11.34	14,922	25.84	11.55
	II	6,383	18.39	5.91	5,872	18.13	6.04
	III	7,160	20.04	6.64	6,968	19.88	6.78
	IV	10,261	18.36	7.26	10,134	18.19	7.27
	V	10,678	20.80	6.84	10,368	20.48	6.99
Mathematics	I	14,673	23.77	10.98	14,866	23.74	11.16
	II	6,381	20.19	6.22	5,864	20.01	6.36
	III	7,142	20.01	6.27	6,952	19.78	6.38
	IV	10,241	18.46	6.49	10,103	18.26	6.60
	V	10,644	20.31	7.32	10,324	20.12	7.34
Science	I	3,724	24.39	11.36	3,800	24.47	11.93
	III	3,446	21.02	5.84	3,551	20.91	5.69
	IV	3,275	21.51	5.99	3,290	21.32	6.40
	V	3,435	20.20	5.94	3,450	20.18	5.96

## 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 (Dragow, 1988). The ETS proprietary software Generalized Analysis System (GENASYS) estimates the value of  $\beta$  for each item using maximum likelihood. In turn, it uses this estimate of  $\beta$  to compute the polyserial correlation from the following formula:

$$r_{polyreg} = \frac{\hat{\beta}s_{tot}}{\sqrt{\hat{\beta}^2 s_{tot}^2 + 1}} \quad (8.1)$$

where,

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

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

$\beta$  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,  $\beta$ . For a polytomously scored item, there are  $k-1$  regressions, where  $k$  is the number of score points on the item. Beta ( $\beta$ ) 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 0.60
- 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
English– Language Arts	I	8	14,922	3.24	0.79	2.85	0.67	3.60	0.84
	II	8	5,872	2.27	0.74	1.87	0.62	3.56	0.85
	III	8	6,968	2.49	0.79	2.23	0.66	3.28	0.87
	IV	8	10,134	2.28	0.79	1.68	0.61	2.50	0.84
	V	8	10,368	2.57	0.81	2.09	0.69	3.09	0.86
Mathematics	I	8	14,866	2.98	0.77	2.64	0.71	3.34	0.82
	II	8	5,864	2.51	0.76	1.25	0.65	3.21	0.86
	III	8	6,952	2.48	0.73	2.02	0.55	3.11	0.84
	IV	8	10,103	2.30	0.75	1.46	0.64	2.89	0.87
	V	8	10,324	2.53	0.78	2.09	0.71	2.88	0.84
Science	I	8	3,800	3.07	0.81	2.57	0.75	3.28	0.84
	III	8	3,551	2.63	0.73	2.21	0.65	3.17	0.79
	IV	8	3,290	2.68	0.75	2.20	0.70	2.94	0.81
	V	8	3,450	2.53	0.76	1.94	0.66	3.36	0.82

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:

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

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 was reported in equation 8.3. The mathematic form of the SEM is as follows:

$$s_e = s_t \sqrt{1 - \alpha} \quad (8.3)$$

where,

$\alpha$  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 particularly useful in determining the confidence interval (CI) that captures an examinee's true score. Assuming that measurement error is normally distributed, it can be said that upon infinite replications of the testing occasion, approximately 95 percent of the CIs of  $\pm 1.96$  SEM around the observed score would contain an examinee's true score (Crocker & Algina, 1986). For example, if an examinee's observed score on a given test equals 15 points, and the SEM equals 1.92, one can be 95 percent confident that the

examinee's true score lies between 11 and 19 points ( $15 \pm 3.76$  rounded to the nearest integer).

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

**Table 8.3 Reliabilities and SEMs for the CAPA**

Content Area	Level	No. of Items	No. of Examinees	Reliab.	Scale Score			Raw Score		
					Mean	S.D.	SEM	Mean	S.D.	SEM
English– Language Arts	I	8	14,922	0.88	41.45	10.85	3.68	25.84	11.55	3.92
	II	8	5,872	0.84	38.29	6.24	2.47	18.13	6.04	2.39
	III	8	6,968	0.89	39.44	5.98	2.01	19.88	6.78	2.27
	IV	8	10,134	0.89	38.95	8.18	2.74	18.19	7.27	2.43
	V	8	10,368	0.90	38.66	6.56	2.08	20.48	6.99	2.22
Mathematics	I	8	14,866	0.86	36.61	9.52	3.50	23.74	11.16	4.11
	II	8	5,864	0.85	37.25	8.66	3.34	20.01	6.36	2.45
	III	8	6,952	0.84	36.26	5.86	2.37	19.78	6.38	2.57
	IV	8	10,103	0.84	36.55	7.67	3.05	18.26	6.60	2.63
	V	8	10,324	0.87	37.27	7.95	2.86	20.12	7.34	2.64
Science	I	8	3,800	0.90	37.61	11.14	3.58	24.47	11.93	3.83
	III	8	3,551	0.84	36.09	4.65	1.85	20.91	5.69	2.27
	IV	8	3,290	0.86	35.73	5.69	2.10	21.32	6.40	2.35
	V	8	3,450	0.85	35.86	5.12	2.01	20.18	5.96	2.33

### 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. The 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. Finally, results based on samples that contain 50 or fewer examinees should be interpreted with caution due to small sample sizes.

### Conditional Standard Errors of Measurement

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

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

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

where,

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

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

The statistic is multiplied by  $a$ , where  $a$  is the original scaling factor needed to transform theta to the scale-score metric. The value of  $a$  varies by 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.1 through Table 8.D.14 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 at 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 ETS-proprietary computer program RELCLASS-COMP (Version 4.14).

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

Decision consistency describes the extent to which examinees are classified in the same way as they would be on the basis of a single form of a test other than the one for which data are available. Decision consistency answers the following question: What is the agreement between the classifications based on two nonoverlapping, equally difficult forms of the test? RELCLASS-COMP also estimates decision consistency using an estimated multivariate distribution of reported classifications on the current form of the exam 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 103.

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

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

## Validity Evidence

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

Validation is the process of accumulating evidence to support each proposed score interpretation or use. It involves more than a single study or gathering of one particular kind of evidence. Validation involves multiple investigations and various kinds of evidence (AERA, APA, & NCME, 1999; Cronbach, 1971; ETS, 2002; Kane, 2006). The process begins with test design and continues through the entire assessment process, including task 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 (USDOE) for the peer review of testing programs administered by states in response to the Elementary and Secondary Education Act (USDOE, 2001).

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

### **Purposes of the CAPA**

As mentioned in Chapter 1, CAPA results for ELA and mathematics in 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 are 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 developed all CAPA tasks to conform to the SBE-approved content standards and test blueprints.

### **Interpretations and Uses of the Scores Generated**

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

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

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

3. Evaluating school programs; and
4. Providing data for federal accountability programs for schools (grade ten ELA and mathematics only).

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 Smarter Balanced for ELA and mathematics and the CST or CMA for Science with or without universal tools, designated supports, and accommodations. Participation in the CAPA and eligibility are determined by a student's IEP team. Only those students whose parents/guardians have submitted written requests to exempt them from CAASPP System 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 *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 17. Once the tasks were developed and field-tested, ETS selected 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. Test blueprints were also reviewed and approved by the CDE and presented to the SBE for adoption. There have been no recent changes in the blueprints for the 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 17.

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

**Form construction process**—For each test, the content standards, blueprints, and test specifications were used as the basis for choosing tasks. Additional targets for item difficulty and discrimination that were used for test construction were defined in light of what are desirable statistical characteristics in test 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), was responsible for reviewing and approving the achievement tests to be used statewide for the testing of students in California public schools, grades two through eleven. More information about the SPAR is given in Chapter 3, starting on page 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 (HumRRO, 2007). 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.

### **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 *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 provided evidence of the degree to which a score interpretation or use was valid for individuals who differ in particular demographic characteristics. For the CAPA, DIF analyses were performed after the test forms' original administration in 2013 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 of the *2013 CAPA Technical Report*, which is the report for the year each form was administered originally. The report is linked on the CDE's Technical Reports and Studies Web page at <http://www.cde.ca.gov/ta/tg/sr/technicalrpts.asp>.

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-area Test Scores**

To the degree that students' content-area test scores correlate as expected, evidence of the validity in regarding those scores as measures of the intended constructs is provided. Table 8.4 provides the correlations between scores on the 2014 CAPA content-area tests and the numbers of students on which these correlations were based. Sample sizes for individual tests are shown in bold font 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 test score 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 while the correlations are reported only for samples that comprise 11 or more examinees, results based on samples that contain 50 or fewer examinees should be interpreted with caution due to small sample sizes. Correlations between scores on any two content-area tests where 10 or fewer examinees with valid scores are expressed as hyphens. Correlations between scores on 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	<b>14,922</b>	0.81	0.81
	Mathematics	14,860	<b>14,866</b>	0.80
	Science	3,794	3,792	<b>3,800</b>
II	ELA	<b>5,872</b>	0.71	N/A
	Mathematics	5,855	<b>5,864</b>	N/A
	Science	N/A	N/A	N/A
III	ELA	<b>6,968</b>	0.76	0.74
	Mathematics	6,949	<b>6,952</b>	0.72
	Science	3,538	3,536	<b>3,551</b>
IV	ELA	<b>10,134</b>	0.76	0.71
	Mathematics	10,095	<b>10,103</b>	0.70
	Science	3,272	3,269	<b>3,290</b>
V	ELA	<b>10,368</b>	0.76	0.74
	Mathematics	10,306	<b>10,324</b>	0.71
	Science	3,432	3,433	<b>3,450</b>

### 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. Across all CAPA levels for ELA, mathematics, and science, the percentage of students for whom the raters were in exact agreement ranged from 91 percent to 98 percent. The results are presented in Table 8.C.36 to Table 8.C.40.

### 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 the 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 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 0.84 to 0.90.

**Subgroup reliabilities**—The reliabilities of the operational CAPA are 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 through 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 the intended use of the scores. The CDE and ETS are in the process of determining what kinds of information can be gathered to assess the consequences of the administration of the CAPA.

## IRT Analyses

### Post-Equating

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

The procedures used for post-equating the CAPA prior to 2014 involved three steps: task calibration, task parameter scaling, and production of raw-score-to-scale-score conversions using the scaled task parameters. ETS used GENASYS for the IRT item calibration and equating work. The IRT model used to calibrate the CAPA test tasks was the one-parameter partial credit (1PPC) model, a more restrictive version of the generalized partial-credit model (Muraki, 1992), in which all tasks were assumed to be equally discriminating. This model stated that the probability that an examinee with ability  $\theta$  will perform in the  $k$ th category of  $m_j$  ordered score categories of task  $j$  can be expressed as:

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

where,

$m_j$  is the number of possible score categories ( $c=1 \dots m_j$ ) 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.

### Pre-Equating

During the 2014 administration, because intact test forms from the 2013 administration were reused, the conversion tables from the previous administration when the forms were originally used are directly applied to the 2014 operational scoring.

Descriptions of IRT analyses such as the model-data fit analyses can be found in Chapter 8 of the original-year (2013) technical report; the results of the IRT analyses are presented in Appendix 8.D of the original-year (2013) technical report. The *2013 CAPA Technical Report*, which is the report for the year each form was administered originally, is linked on the CDE's Technical Reports and Studies Web page at <http://www.cde.ca.gov/ta/tg/sr/technicalrpts.asp>.

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

For the post-equating procedure prior to the 2014 administration, once the IRT  $b$ -values were placed on the item bank scale, analyses were performed to assess the overall test difficulty and the distribution of tasks in a particular range of item difficulty.

During the 2014 administration, neither IRT calibrations nor scaling are implemented. The summaries of  $b$ -values can be found in Appendix D of the *2013 CAPA Technical Report*, which is the report for the year each form was administered originally.

### Equating Results

During the 2014 administration, for the reused intact forms, the conversion tables from their original administrations (2013) are directly applied to the current administration.

Complete raw-score-to-scale-score conversion tables for the CAPA administered in 2014 based on P1 data are presented in Table 8.D.1 through Table 8.D.14, starting on page 127. The raw scores and corresponding transformed scale scores are listed in those tables. For all of the 2014 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 who differ in their demographic characteristics.

Prior to the 2014 administration, DIF analyses were performed based on the final item analyses (FIA) sample and were performed on all operational tasks and all field-test tasks for which sufficient student samples were available. DIF analyses are not implemented during the 2014 administration because forms are reused and all tasks were evaluated for DIF during the previous administration when the intact forms were originally used. These DIF results can be found in Appendix E of the *2013 CAPA Technical Report*, which is the report for the year the form was administered originally.

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

The sample size requirements for the DIF analyses were 100 in the focal group and 400 in the combined focal and reference groups. These sample sizes were based on standard operating procedures with respect to DIF analyses at ETS.

DIF analyses of the polytomously scored CAPA tasks were completed using two procedures. The first was 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 was 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  $\alpha$  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 was 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:

$$SMD = \sum_m w_m (E_{fm} - E_{rm}) / \sum_m w_m \quad (8.5)$$

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 $0.17 \leq  SMD/SD  < 0.25$ .
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 8.5. The categories have been used by all ETS testing programs for more than 15 years.

**Table 8.5 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 8.5, 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.

Table 8.6 lists specific subgroups that were used for DIF analyses for the CAPA including primary disability. Table 8.E.1 to Table 8.E.5, starting on page 141 in Appendix E, show the sample size for disability groups within CAPA test level and content area.

Table 8.6 Subgroup Classification for DIF Analyses

DIF Type	Reference Group		Focal Group
	Male	Female	
<b>Gender</b>			
<b>Race/Ethnicity</b>	White		<ul style="list-style-type: none"> <li>• African American</li> <li>• American Indian</li> <li>• Asian</li> <li>• Combined Asian Group (Asian/Pacific Islander/Filipino)</li> <li>• Filipino</li> <li>• Hispanic/Latin American</li> <li>• Pacific Islander</li> </ul>
<b>Disability</b>	Mental Retardation/ Intellectual Disability (MR/ID)		<ul style="list-style-type: none"> <li>• Autism</li> <li>• Deaf-Blindness</li> <li>• Deafness</li> <li>• Emotional Disturbance</li> <li>• Hard of Hearing</li> <li>• Multiple Disabilities</li> <li>• Orthopedic Impairment</li> <li>• Other Health Impairment</li> <li>• Specific Learning Disability</li> <li>• Speech or Language Impairment</li> <li>• Traumatic Brain Injury</li> <li>• Visual Impairment</li> </ul>

## References

- AERA, APA, & NCME 1999. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- California Department of Education. (2013). *California Alternate Performance Assessment technical report, spring 2013 administration*. Sacramento, CA. Retrieved from <http://www.cde.ca.gov/ta/tg/sr/documents/capa13techrpt.pdf>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: Holt.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 292–334.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443–507). Washington, 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, 107–114.
- 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.
- Dragow 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.
- HumRRO. (2007). *Independent evaluation of the alignment of the California Standards Tests (CSTs) and the California Alternate Performance Assessment (CAPA)*. Alexandria, VA: Author. Retrieved from <http://www.cde.ca.gov/ta/tg/sr/documents/alignmentreport.pdf>
- 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., & Bock, R. D. (1995). *PARSCALE: Parameter scaling of rating data* (Computer software, Version 2.2). Chicago, IL: Scientific Software.
- Stocking, M. L., and Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*, pp. 201–10.
- 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—Current Year (2014) and Original Year of Administration (2013)**

**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

		2013			2014			
Form	Task Position	AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	1	3.30	.81		1	3.33	.82	
Operational	3	3.31	.79		3	3.29	.80	
Operational	4	3.59	.83		4	3.60	.84	
Operational	6	2.97	.77		6	3.00	.78	
Operational	7	3.64	.66		7	3.58	.67	
Operational	9	3.15	.81		9	3.13	.82	
Operational	10	3.16	.79		10	3.16	.79	
Operational	12	2.87	.76		12	2.85	.76	

**Table 8.A.2 AIS and Polyserial Correlation: Level II, ELA—Current Year (2014) and Original Year of Administration (2013)**

**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

		2013			2014			
Form	Task Position	AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	1	2.03	.68		1	2.02	.70	
Operational	3	3.57	.70	H	3	3.56	.71	H
Operational	4	2.00	.83		4	1.93	.83	
Operational	6	2.15	.83		6	2.13	.85	
Operational	7	2.46	.82		7	2.43	.82	
Operational	9	2.43	.68		9	2.38	.70	
Operational	10	1.88	.71		10	1.87	.72	
Operational	12	1.87	.63		12	1.87	.62	

**Table 8.A.3 AIS and Polyserial Correlation: Level III, ELA—Current Year (2014) and Original Year of Administration (2013)**

**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

		2013			2014			
Form	Task Position	AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	1	2.40	.85		1	2.38	.86	
Operational	3	2.62	.85		3	2.62	.85	
Operational	4	2.47	.85		4	2.43	.86	
Operational	6	2.27	.67		6	2.27	.69	
Operational	7	3.33	.79	H	7	3.28	.79	H
Operational	9	2.28	.69		9	2.29	.71	
Operational	10	2.25	.87		10	2.23	.87	
Operational	12	2.47	.66		12	2.45	.66	

**Table 8.A.4 AIS and Polyserial Correlation: Level IV, ELA—Current Year (2014) and Original Year of Administration (2013)**

**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

		2013			2014			
Form	Task Position	AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	1	2.27	.61		1	2.29	.61	
Operational	3	2.48	.81		3	2.46	.81	
Operational	4	2.53	.82		4	2.50	.82	
Operational	6	2.56	.84		6	2.49	.84	
Operational	7	1.68	.77		7	1.68	.76	
Operational	9	2.25	.81		9	2.24	.81	
Operational	10	2.40	.82		10	2.36	.81	
Operational	12	2.20	.83		12	2.23	.83	

**Table 8.A.5 AIS and Polyserial Correlation: Level V, ELA—Current Year (2014) and Original Year of Administration (2013)**

**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

		2013			2014			
Form	Task Position	AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	1	2.25	.82		1	2.24	.81	
Operational	3	3.09	.69		3	3.05	.69	
Operational	4	2.96	.81		4	2.91	.83	
Operational	6	2.52	.83		6	2.46	.84	
Operational	7	3.13	.79		7	3.09	.79	
Operational	9	2.52	.85		9	2.48	.86	
Operational	10	2.13	.76		10	2.09	.77	
Operational	12	2.26	.85		12	2.23	.85	

**Table 8.A.6 AIS and Polyserial Correlation: Level I, Mathematics—Current Year (2014) and Original Year of Administration (2013)**

**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

		2013			2014			
Form	Task Position	AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	13	3.33	.80		13	3.34	.81	
Operational	15	2.67	.79		15	2.71	.80	
Operational	16	2.62	.71		16	2.64	.71	
Operational	18	2.82	.71		18	2.82	.72	
Operational	19	2.92	.74		19	3.00	.75	
Operational	21	3.31	.82		21	3.33	.82	
Operational	22	2.68	.77		22	2.69	.77	
Operational	24	3.36	.80		24	3.34	.80	

**Table 8.A.7 AIS and Polyserial Correlation: Level II, Mathematics—Current Year (2014) and Original Year of Administration (2013)****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

		2013			2014			
Form	Task Position	AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	13	2.59	.75		13	2.58	.76	
Operational	15	2.98	.83		15	2.96	.84	
Operational	16	3.23	.73	H	16	3.21	.77	H
Operational	18	2.12	.80		18	2.12	.80	
Operational	19	2.86	.73		19	2.87	.75	
Operational	21	2.34	.85		21	2.31	.86	
Operational	22	1.25	.66		22	1.25	.65	
Operational	24	2.79	.68		24	2.77	.68	

**Table 8.A.8 AIS and Polyserial Correlation: Level III, Mathematics—Current Year (2014) and Original Year of Administration (2013)****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

		2013			2014			
Form	Task Position	AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	13	3.12	.81		13	3.11	.80	
Operational	15	2.46	.54	R	15	2.37	.55	R
Operational	16	2.07	.63		16	2.02	.64	
Operational	18	2.84	.80		18	2.81	.81	
Operational	19	2.34	.84		19	2.35	.84	
Operational	21	2.80	.63		21	2.77	.64	
Operational	22	2.09	.82		22	2.10	.83	
Operational	24	2.32	.73		24	2.31	.75	

**Table 8.A.9 AIS and Polyserial Correlation: Level IV, Mathematics—Current Year (2014) and Original Year of Administration (2013)**

**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

		2013			2014			
Form	Task Position	AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	13	1.79	.75		13	1.78	.75	
Operational	15	1.46	.67		15	1.46	.68	
Operational	16	2.51	.86		16	2.45	.86	
Operational	18	2.91	.81		18	2.88	.81	
Operational	19	1.54	.68		19	1.56	.69	
Operational	21	2.90	.66		21	2.89	.67	
Operational	22	2.66	.60		22	2.66	.64	
Operational	24	2.72	.88		24	2.69	.87	

**Table 8.A.10 AIS and Polyserial Correlation: Level V, Mathematics—Current Year (2014) and Original Year of Administration (2013)**

**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

		2013			2014			
Form	Task Position	AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	13	2.11	.75		13	2.09	.75	
Operational	15	2.82	.73		15	2.80	.74	
Operational	16	2.62	.74		16	2.60	.75	
Operational	18	2.76	.71		18	2.84	.71	
Operational	19	2.24	.78		19	2.23	.79	
Operational	21	2.21	.80		21	2.15	.81	
Operational	22	2.93	.81		22	2.88	.81	
Operational	24	2.71	.84		24	2.65	.84	

**Table 8.A.11 AIS and Polyserial Correlation: Level I, Science—Current Year (2014) and Original Year of Administration (2013)**

**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

		2013			2014			
Form	Task Position	AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	25	3.16	.82		25	3.18	.84	
Operational	27	3.11	.79		27	3.14	.81	
Operational	28	3.05	.76		28	3.12	.77	
Operational	30	3.01	.80		30	3.03	.82	
Operational	31	3.26	.82		31	3.28	.84	
Operational	33	2.58	.77		33	2.57	.79	
Operational	34	3.07	.73		34	3.08	.75	
Operational	36	3.10	.80		36	3.16	.82	

**Table 8.A.12 AIS and Polyserial Correlation: Level III, Science—Current Year (2014) and Original Year of Administration (2013)**

**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

		2013			2014			
Form	Task Position	AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	25	2.52	.81		25	2.52	.79	
Operational	27	2.55	.64		27	2.55	.65	
Operational	28	2.22	.70		28	2.21	.70	
Operational	30	2.50	.78		30	2.50	.76	
Operational	31	2.78	.74		31	2.74	.72	
Operational	33	3.00	.75		33	2.95	.75	
Operational	34	2.36	.74		34	2.36	.71	
Operational	36	3.12	.76		36	3.17	.73	

**Table 8.A.13 AIS and Polyserial Correlation: Level IV, Science—Current Year (2014) and Original Year of Administration (2013)**

**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

Form	Task Position	2013			2014			
		AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	25	2.52	.71		25	2.54	.71	
Operational	27	2.93	.76		27	2.91	.77	
Operational	28	2.51	.67		28	2.47	.70	
Operational	30	2.88	.70		30	2.85	.72	
Operational	31	2.17	.78		31	2.20	.78	
Operational	33	2.92	.74		33	2.92	.78	
Operational	34	2.97	.80		34	2.94	.81	

**Table 8.A.14 AIS and Polyserial Correlation: Level V, Science—Current Year (2014) and Original Year of Administration (2013)**

**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

Form	Task Position	2013			2014			
		AIS	Polyserial	Flag	Task Position	AIS	Polyserial	Flag
Operational	25	1.96	.67		25	1.94	.66	
Operational	27	3.24	.79	H	27	3.28	.78	H
Operational	28	2.16	.78		28	2.18	.79	
Operational	30	2.16	.73		30	2.14	.72	
Operational	31	2.48	.74		31	2.43	.74	
Operational	33	2.87	.80		33	2.88	.81	
Operational	34	3.38	.79	H	34	3.36	.82	H
Operational	36	1.99	.73		36	2.04	.74	

**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												
I	0	1,646	10.75	1,677	10.95	1,593	10.40	1,847	12.06	1,960	12.80	1,751	11.43	1,750	11.42	1,976	12.90
	1	3,672	23.97	3,646	23.80	2,835	18.51	4,293	28.03	2,336	15.25	4,251	27.75	4,134	26.99	4,601	30.04
	2	568	3.71	638	4.17	561	3.66	1,105	7.21	479	3.13	607	3.96	619	4.04	917	5.99
	3	639	4.17	708	4.62	549	3.58	614	4.01	609	3.98	673	4.39	695	4.54	958	6.25
	4	1,013	6.61	1,296	8.46	983	6.42	999	6.52	1,598	10.43	1,112	7.26	1,162	7.59	1,552	10.13
	5	7,780	50.79	7,353	48.00	8,797	57.43	6,460	42.17	8,336	54.42	6,924	45.20	6,958	45.42	5,314	34.69
II	0	377	6.04	295	4.73	679	10.88	609	9.76	671	10.75	336	5.38	450	7.21	432	6.92
	1	2,140	34.28	203	3.25	2,077	33.27	1,709	27.38	1,435	22.99	1,405	22.51	2,307	36.96	2,186	35.02
	2	1,685	26.99	438	7.02	1,236	19.80	1,288	20.63	915	14.66	1,798	28.80	2,013	32.25	2,277	36.48
	3	1,385	22.19	845	13.54	1,845	29.56	1,974	31.62	1,416	22.69	1,407	22.54	935	14.98	871	13.95
	4	655	10.49	4,461	71.47	405	6.49	662	10.61	1,805	28.92	1,296	20.76	537	8.60	476	7.63
III	0	660	8.95	527	7.14	627	8.50	392	5.31	415	5.63	415	5.63	648	8.78	441	5.98
	1	1,928	26.14	952	12.90	1,603	21.73	1,261	17.09	581	7.88	921	12.48	2,282	30.93	1,680	22.77
	2	1,242	16.84	1,808	24.51	1,359	18.42	3,073	41.66	952	12.90	3,438	46.60	1,218	16.51	2,250	30.50
	3	1,642	22.26	2,277	30.87	2,159	29.27	1,853	25.12	924	12.53	1,889	25.61	1,815	24.60	970	13.15
	4	1,905	25.82	1,813	24.58	1,629	22.08	798	10.82	4,505	61.07	714	9.68	1,414	19.17	2,036	27.60
IV	0	613	5.75	807	7.57	674	6.32	837	7.85	1,352	12.67	670	6.28	934	8.76	923	8.65
	1	3,529	33.08	2,174	20.38	1,952	18.30	2,451	22.98	5,412	50.74	4,159	38.99	3,040	28.50	3,387	31.75
	2	2,028	19.01	2,792	26.17	2,237	20.97	1,938	18.17	1,373	12.87	1,073	10.06	1,757	16.47	1,789	16.77
	3	2,075	19.45	2,185	20.48	4,087	38.31	2,672	25.05	1,016	9.52	2,458	23.04	2,221	20.82	2,561	24.01
	4	2,422	22.71	2,709	25.40	1,717	16.10	2,769	25.96	1,514	14.19	2,307	21.63	2,715	25.45	2,007	18.82
V	0	865	7.89	778	7.10	935	8.53	992	9.05	864	7.88	1,035	9.44	882	8.04	1,125	10.26
	1	4,064	37.07	662	6.04	1,436	13.10	2,030	18.52	777	7.09	2,215	20.20	4,487	40.92	2,780	25.36
	2	984	8.97	1,342	12.24	1,314	11.98	2,208	20.14	1,709	15.59	1,947	17.76	1,679	15.31	2,131	19.44
	3	3,013	27.48	4,452	40.61	2,953	26.93	3,801	34.67	2,635	24.03	3,511	32.02	1,871	17.06	3,689	33.65
	4	2,038	18.59	3,730	34.02	4,326	39.46	1,933	17.63	4,979	45.41	2,256	20.58	2,045	18.65	1,239	11.30

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

Math Score on Level	Task	1		2		3		4		5		6		7		8	
		Count	Percent														
I	0	1,746	11.40	2,367	15.45	1,927	12.58	1,624	10.60	1,859	12.14	1,690	11.03	2,097	13.69	1,964	12.82
	1	3,453	22.54	4,751	31.02	5,584	36.45	5,538	36.15	4,587	29.95	3,799	24.80	5,310	34.67	3,326	21.71
	2	569	3.71	947	6.18	751	4.90	639	4.17	572	3.73	525	3.43	768	5.01	632	4.13
	3	728	4.75	838	5.47	956	6.24	777	5.07	784	5.12	575	3.75	838	5.47	666	4.35
	4	1,267	8.27	1,117	7.29	1,451	9.47	1,159	7.57	1,252	8.17	907	5.92	1,138	7.43	1,259	8.22
II	0	378	6.06	265	4.25	266	4.26	396	6.34	318	5.09	402	6.44	662	10.61	381	6.10
	1	844	13.52	1,171	18.76	367	5.88	3,070	49.18	905	14.50	2,356	37.74	4,368	69.98	788	12.62
	2	690	11.05	681	10.91	653	10.46	461	7.39	1,005	16.10	766	12.27	663	10.62	1,496	23.97
	3	3,915	62.72	1,123	17.99	2,089	33.47	419	6.71	1,651	26.45	738	11.82	301	4.82	1,523	24.40
	4	415	6.65	3,002	48.09	2,867	45.93	1,896	30.37	2,363	37.86	1,980	31.72	248	3.97	2,054	32.91
III	0	370	5.02	376	5.10	442	5.99	410	5.56	500	6.78	371	5.03	548	7.43	446	6.05
	1	1,142	15.48	1,460	19.79	2,843	38.54	1,758	23.83	2,813	38.13	697	9.45	3,179	43.09	1,409	19.10
	2	906	12.28	2,682	36.36	1,692	22.94	982	13.31	835	11.32	1,193	16.17	834	11.31	2,007	27.21
	3	720	9.76	1,409	19.10	1,480	20.06	691	9.37	685	9.29	3,882	52.62	1,221	16.55	3,104	42.08
	4	4,239	57.46	1,450	19.66	920	12.47	3,536	47.93	2,544	34.49	1,234	16.73	1,595	21.62	411	5.57
IV	0	1,084	10.16	914	8.57	749	7.02	734	6.88	696	6.52	645	6.05	628	5.89	945	8.86
	1	5,617	52.66	5,642	52.89	4,076	38.21	2,490	23.34	6,966	65.30	1,283	12.03	1,689	15.83	3,280	30.75
	2	1,255	11.77	3,359	31.49	709	6.65	793	7.43	1,135	10.64	2,106	19.74	2,366	22.18	551	5.17
	3	824	7.72	455	4.27	1,122	10.52	1,416	13.27	756	7.09	2,666	24.99	3,268	30.64	902	8.46
	4	1,887	17.69	297	2.78	4,011	37.60	5,234	49.07	1,114	10.44	3,967	37.19	2,716	25.46	4,989	46.77
V	0	927	8.45	780	7.11	936	8.54	916	8.35	834	7.61	1,006	9.18	941	8.58	977	8.91
	1	4,544	41.44	1,130	10.31	2,676	24.41	2,892	26.38	3,200	29.19	4,323	39.43	2,036	18.57	2,866	26.14
	2	1,753	15.99	3,804	34.70	1,956	17.84	628	5.73	2,894	26.40	1,529	13.95	1,492	13.61	1,290	11.77
	3	1,396	12.73	822	7.50	1,513	13.80	1,013	9.24	2,100	19.15	1,628	14.85	1,245	11.36	1,712	15.61
	4	2,344	21.38	4,428	40.39	3,883	35.42	5,515	50.30	1,936	17.66	2,478	22.60	5,250	47.88	4,119	37.57

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

Science Score on		1		2		3		4		5		6		7		8	
Level	Task	Count	Percent														
I	0	814	17.47	835	17.92	801	17.19	840	18.03	792	17.00	939	20.15	776	16.65	880	18.88
	1	1,146	24.59	1,125	24.14	1,225	26.29	1,258	27.00	1,089	23.37	1,569	33.67	1,276	27.38	1,044	22.40
	2	175	3.76	202	4.33	167	3.58	187	4.01	153	3.28	204	4.38	179	3.84	258	5.54
	3	194	4.16	228	4.89	223	4.79	219	4.70	187	4.01	285	6.12	249	5.34	239	5.13
	4	322	6.91	370	7.94	359	7.70	360	7.73	266	5.71	353	7.58	341	7.32	308	6.61
	5	2,009	43.11	1,900	40.77	1,885	40.45	1,796	38.54	2,173	46.63	1,310	28.11	1,839	39.46	1,931	41.44
III	0	295	7.58	265	6.81	326	8.37	308	7.91	282	7.24	284	7.29	305	7.83	305	7.83
	1	670	17.21	468	12.02	895	22.98	497	12.76	477	12.25	413	10.61	729	18.72	252	6.47
	2	988	25.37	1,478	37.96	1,212	31.12	1,327	34.08	963	24.73	715	18.36	1,152	29.58	377	9.68
	3	1,184	30.41	803	20.62	1,055	27.09	1,038	26.66	1,059	27.20	1,014	26.04	1,241	31.87	1,375	35.31
	4	757	19.44	880	22.60	406	10.43	724	18.59	1,113	28.58	1,468	37.70	467	11.99	1,585	40.70
IV	0	316	8.41	346	9.21	379	10.09	381	10.14	473	12.59	355	9.45	368	9.80	380	10.11
	1	470	12.51	448	11.92	678	18.05	559	14.88	1,099	29.25	245	6.52	400	10.65	618	16.45
	2	1,213	32.29	772	20.55	925	24.62	625	16.64	845	22.49	656	17.46	704	18.74	867	23.08
	3	1,095	29.15	666	17.73	1,095	29.15	811	21.59	522	13.89	1,481	39.42	798	21.24	1,011	26.91
	4	663	17.65	1,525	40.59	680	18.10	1,381	36.76	818	21.77	1,020	27.15	1,487	39.58	881	23.45
V	0	434	10.68	414	10.19	509	12.53	484	11.92	447	11.00	440	10.83	435	10.71	526	12.95
	1	1,358	33.43	241	5.93	1,164	28.66	1,194	29.39	767	18.88	383	9.43	236	5.81	1,303	32.08
	2	1,258	30.97	375	9.23	1,007	24.79	1,310	32.25	1,108	27.28	706	17.38	380	9.35	1,090	26.83
	3	718	17.68	1,034	25.46	686	16.89	250	6.15	958	23.58	1,328	32.69	658	16.20	593	14.60
	4	294	7.24	1,998	49.19	696	17.13	824	20.29	782	19.25	1,205	29.67	2,353	57.93	550	13.54

## Appendix 8.B—Reliability Analyses

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

**Table 8.B.1 Reliabilities and SEMs by Gender**

Content Area	Level	Male			Female			Unknown Gender		
		N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM
English– Language Arts	I	9,470	0.88	3.95	5,357	0.90	3.85	95	0.91	3.69
	II	4,086	0.85	2.38	1,729	0.83	2.39	57	0.85	2.35
	III	4,743	0.89	2.28	2,191	0.88	2.26	34	0.90	2.48
	IV	6,670	0.88	2.46	3,406	0.90	2.36	58	0.89	2.42
	V	6,799	0.90	2.23	3,518	0.90	2.19	51	0.89	2.24
Mathematics	I	9,435	0.86	4.13	5,335	0.88	4.06	96	0.89	3.93
	II	4,084	0.85	2.45	1,723	0.85	2.44	57	0.85	2.46
	III	4,732	0.84	2.58	2,188	0.83	2.54	32	0.72	2.88
	IV	6,650	0.84	2.64	3,395	0.84	2.60	58	0.84	2.53
	V	6,770	0.87	2.64	3,504	0.86	2.64	50	0.88	2.61
Science	I	2,352	0.89	3.88	1,431	0.91	3.77	17	0.96	3.07
	III	2,422	0.85	2.27	1,114	0.83	2.25	15	0.84	2.28
	IV	2,130	0.87	2.36	1,153	0.86	2.34	7	–	–
	V	2,291	0.86	2.32	1,148	0.82	2.35	11	0.93	2.17

Table 8.B.2 Reliabilities and SEMs by Primary Ethnicity

Content Area	Level	American Indian			Asian			Pacific Islander			Filipino		
		N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM
English– Language Arts	I	108	0.88	3.92	1,203	0.86	4.18	76	0.89	3.89	429	0.85	4.12
	II	40	0.85	2.22	383	0.84	2.39	28	0.80	2.16	177	0.84	2.40
	III	47	0.82	2.28	446	0.89	2.33	24	0.87	2.57	224	0.86	2.40
	IV	96	0.88	2.44	658	0.89	2.42	57	0.87	2.56	296	0.88	2.41
	V	98	0.88	2.16	674	0.91	2.24	57	0.90	2.17	343	0.91	2.27
Mathematics	I	106	0.86	4.12	1,198	0.83	4.30	75	0.84	4.26	429	0.84	4.21
	II	40	0.86	2.24	383	0.86	2.50	27	0.80	2.52	179	0.86	2.43
	III	47	0.82	2.41	445	0.85	2.59	23	0.81	2.63	224	0.82	2.61
	IV	96	0.81	2.74	654	0.86	2.60	57	0.86	2.55	295	0.86	2.60
	V	97	0.86	2.66	674	0.89	2.59	57	0.88	2.65	340	0.87	2.69
Science	I	27	0.88	3.81	285	0.86	4.06	17	0.90	3.78	111	0.90	3.75
	III	23	0.78	2.06	232	0.86	2.34	12	0.36	2.58	116	0.84	2.37
	IV	31	0.88	2.42	207	0.84	2.51	20	0.86	2.42	109	0.89	2.26
	V	38	0.87	2.32	237	0.87	2.32	19	0.85	2.35	116	0.85	2.35
Content Area	Level	Hispanic			African American			White			Unknown Ethnicity		
		N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM
English– Language Arts	I	7,983	0.90	3.81	1,252	0.89	3.85	3,208	0.87	4.04	663	0.86	4.03
	II	3,263	0.84	2.38	470	0.85	2.35	1,230	0.85	2.41	281	0.85	2.39
	III	3,867	0.89	2.25	572	0.89	2.27	1,521	0.89	2.26	267	0.87	2.38
	IV	5,385	0.89	2.38	932	0.88	2.46	2,237	0.89	2.46	473	0.89	2.43
	V	5,145	0.90	2.18	1,094	0.89	2.23	2,552	0.90	2.24	405	0.90	2.24
Mathematics	I	7,956	0.87	4.05	1,242	0.87	4.02	3,198	0.85	4.16	662	0.85	4.16
	II	3,256	0.85	2.44	473	0.84	2.46	1,225	0.86	2.45	281	0.86	2.41
	III	3,861	0.84	2.56	569	0.84	2.58	1,516	0.84	2.57	267	0.80	2.69
	IV	5,371	0.84	2.63	929	0.83	2.63	2,230	0.85	2.64	471	0.86	2.61
	V	5,121	0.87	2.62	1,082	0.86	2.68	2,550	0.87	2.65	403	0.88	2.62
Science	I	2,044	0.91	3.75	335	0.90	3.76	838	0.88	3.99	143	0.88	3.93
	III	1,935	0.83	2.25	296	0.84	2.27	790	0.86	2.25	147	0.79	2.33
	IV	1,732	0.85	2.35	297	0.89	2.27	775	0.87	2.36	119	0.90	2.34
	V	1,697	0.84	2.32	365	0.82	2.35	834	0.84	2.35	144	0.87	2.33

**Table 8.B.3 Reliabilities and SEMs by Primary Ethnicity for Economically Disadvantaged**

Content Area	Level	American Indian			Asian			Pacific Islander			Filipino		
		N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM
English– Language Arts	I	67	0.88	3.76	435	0.87	4.06	40	0.79	4.49	155	0.83	4.06
	II	23	0.82	2.07	156	0.87	2.30	12	0.57	2.30	58	0.82	2.51
	III	32	0.79	2.41	193	0.88	2.33	14	0.87	2.78	71	0.84	2.36
	IV	61	0.89	2.43	260	0.88	2.42	35	0.81	2.61	113	0.88	2.40
	V	66	0.84	2.16	266	0.91	2.23	35	0.89	2.10	122	0.93	2.20
Mathematics	I	66	0.87	4.06	432	0.85	4.23	39	0.78	4.42	155	0.85	4.10
	II	23	0.82	2.35	156	0.86	2.51	11	0.57	2.14	59	0.80	2.43
	III	32	0.85	2.30	192	0.84	2.65	13	0.82	2.59	71	0.73	2.72
	IV	61	0.80	2.74	261	0.86	2.59	35	0.86	2.42	112	0.87	2.53
	V	65	0.85	2.58	266	0.89	2.59	35	0.85	2.81	121	0.89	2.63
Science	I	16	0.84	3.72	112	0.85	4.13	9	–	–	41	0.91	3.61
	III	13	0.79	1.93	95	0.82	2.41	8	–	–	34	0.83	2.45
	IV	17	0.81	2.38	89	0.85	2.43	14	0.86	2.21	33	0.90	2.20
	V	25	0.71	2.28	87	0.89	2.26	15	0.88	2.36	41	0.88	2.26
Content Area	Level	Hispanic			African American			White			Unknown Ethnicity		
		N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM
English– Language Arts	I	5,972	0.90	3.78	776	0.89	3.77	1,104	0.87	3.99	279	0.88	3.89
	II	2,485	0.84	2.37	314	0.85	2.35	517	0.83	2.38	124	0.84	2.42
	III	3,077	0.89	2.24	390	0.88	2.30	646	0.88	2.24	125	0.87	2.38
	IV	4,288	0.89	2.37	624	0.87	2.48	907	0.88	2.45	225	0.89	2.42
	V	4,001	0.90	2.17	749	0.89	2.21	973	0.89	2.22	165	0.88	2.29
Mathematics	I	5,954	0.87	4.04	770	0.86	4.01	1,104	0.86	4.11	278	0.87	3.98
	II	2,479	0.85	2.44	316	0.85	2.45	514	0.85	2.40	124	0.88	2.39
	III	3,073	0.83	2.55	390	0.81	2.63	646	0.84	2.56	125	0.83	2.57
	IV	4,274	0.83	2.62	624	0.82	2.67	904	0.83	2.63	224	0.87	2.62
	V	3,982	0.87	2.62	741	0.86	2.69	973	0.87	2.64	167	0.86	2.73
Science	I	1,559	0.91	3.69	213	0.90	3.75	260	0.88	3.96	64	0.91	3.55
	III	1,565	0.83	2.25	210	0.83	2.24	335	0.83	2.23	69	0.79	2.33
	IV	1,393	0.85	2.34	200	0.86	2.29	324	0.87	2.31	61	0.89	2.37
	V	1,322	0.84	2.32	245	0.82	2.36	321	0.81	2.32	54	0.77	2.53

Table 8.B.4 Reliabilities and SEMs by Primary Ethnicity for Not Economically Disadvantaged

Content Area	Level	American Indian			Asian			Pacific Islander			Filipino		
		N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM
English– Language Arts	I	34	0.86	4.24	694	0.85	4.24	29	0.94	3.43	257	0.86	4.15
	II	11	0.91	2.25	195	0.81	2.46	14	0.86	2.16	111	0.86	2.34
	III	13	0.88	2.06	233	0.89	2.32	8	–	–	141	0.87	2.40
	IV	33	0.83	2.46	366	0.89	2.41	17	0.90	2.60	174	0.89	2.40
	V	31	0.91	2.14	384	0.90	2.25	20	0.92	2.15	209	0.90	2.32
Mathematics	I	34	0.84	4.34	692	0.83	4.32	29	0.90	3.96	257	0.84	4.28
	II	11	0.94	2.03	195	0.85	2.52	14	0.83	2.66	112	0.88	2.42
	III	13	0.74	2.60	233	0.86	2.54	8	–	–	141	0.85	2.56
	IV	33	0.83	2.70	363	0.86	2.61	17	0.86	2.74	174	0.85	2.65
	V	31	0.85	2.84	384	0.89	2.58	20	0.91	2.44	207	0.87	2.72
Science	I	11	0.92	3.75	160	0.87	4.02	7	–	–	66	0.88	3.92
	III	9	–	–	129	0.88	2.30	2	–	–	80	0.84	2.28
	IV	13	0.92	2.36	117	0.84	2.57	5	–	–	72	0.87	2.30
	V	13	0.91	2.42	145	0.87	2.35	4	–	–	74	0.84	2.38
Content Area	Level	Hispanic			African American			White			Unknown Ethnicity		
		N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM
English– Language Arts	I	1,698	0.89	3.95	398	0.86	4.14	1,947	0.87	4.07	255	0.83	4.23
	II	569	0.84	2.43	110	0.87	2.36	623	0.85	2.42	89	0.88	2.33
	III	651	0.90	2.29	153	0.91	2.19	809	0.89	2.28	93	0.89	2.32
	IV	916	0.90	2.41	245	0.90	2.42	1,214	0.89	2.45	143	0.89	2.45
	V	914	0.90	2.22	290	0.88	2.31	1,424	0.90	2.25	138	0.91	2.28
Mathematics	I	1,692	0.88	4.08	395	0.85	4.16	1,938	0.85	4.20	254	0.82	4.38
	II	568	0.85	2.46	110	0.84	2.46	623	0.85	2.50	89	0.87	2.37
	III	649	0.83	2.60	152	0.87	2.51	805	0.85	2.58	93	0.81	2.67
	IV	915	0.85	2.62	244	0.87	2.56	1,212	0.85	2.63	143	0.86	2.62
	V	912	0.87	2.63	288	0.86	2.68	1,424	0.87	2.64	207	0.89	2.54
Science	I	431	0.89	3.94	104	0.89	3.93	550	0.87	4.03	66	0.83	4.39
	III	316	0.84	2.25	77	0.86	2.32	425	0.86	2.28	80	0.78	2.27
	IV	304	0.84	2.42	81	0.92	2.22	422	0.87	2.38	72	0.90	2.42
	V	306	0.83	2.34	109	0.82	2.32	467	0.85	2.37	74	0.90	2.13

**Table 8.B.5 Reliabilities and SEMs by Primary Ethnicity for Unknown Economic Status**

Content Area	Level	American Indian			Asian			Pacific Islander			Filipino		
		N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM
English– Language Arts	I	7	–	–	74	0.84	4.23	7	–	–	17	0.82	4.31
	II	6	–	–	32	0.87	2.39	2	–	–	8	–	–
	III	2	–	–	20	0.87	2.55	2	–	–	12	0.86	2.68
	IV	2	–	–	32	0.92	2.27	5	–	–	9	–	–
	V	1	–	–	24	0.87	2.24	2	–	–	12	0.68	2.18
Mathematics	I	6	–	–	74	0.79	4.48	7	–	–	17	0.84	4.13
	II	6	–	–	32	0.91	2.30	2	–	–	8	–	–
	III	2	–	–	20	0.76	2.60	2	–	–	12	0.79	2.69
	IV	2	–	–	30	0.86	2.60	5	–	–	9	–	–
	V	1	–	–	24	0.82	2.74	2	–	–	12	0.40	2.98
Science	I	11	–	–	13	0.90	4.09	1	–	–	4	–	–
	III	1	–	–	8	–	–	2	–	–	2	–	–
	IV	1	–	–	1	–	–	1	–	–	4	–	–
	V	0	–	–	5	–	–	0	–	–	1	–	–
Content Area	Level	Hispanic			African American			White			Unknown Ethnicity		
		N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM
English– Language Arts	I	313	0.91	3.64	78	0.94	3.17	157	0.88	4.04	129	0.88	3.93
	II	209	0.83	2.35	46	0.83	2.32	90	0.84	2.41	68	0.83	2.42
	III	139	0.88	2.17	29	0.93	2.11	66	0.92	2.12	49	0.86	2.41
	IV	181	0.88	2.42	63	0.88	2.34	116	0.89	2.51	105	0.89	2.38
	V	230	0.90	2.17	55	0.88	2.15	155	0.87	2.23	102	0.92	2.06
Mathematics	I	310	0.87	3.98	77	0.93	3.32	156	0.87	4.06	130	0.87	4.10
	II	209	0.82	2.49	47	0.84	2.47	88	0.85	2.43	68	0.79	2.47
	III	139	0.86	2.53	27	0.89	2.25	65	0.84	2.67	49	0.66	3.00
	IV	182	0.81	2.68	61	0.81	2.59	114	0.84	2.67	104	0.81	2.55
	V	227	0.86	2.70	53	0.82	2.56	153	0.82	2.86	100	0.88	2.57
Science	I	54	0.91	3.81	18	0.88	2.76	28	0.94	3.36	21	0.91	3.68
	III	54	0.87	2.32	9	–	–	30	0.88	2.17	24	0.82	2.37
	IV	35	0.85	2.25	16	0.90	2.35	29	0.90	2.53	20	0.93	2.06
	V	69	0.85	2.29	11	0.77	2.52	46	0.82	2.35	29	0.89	2.40

Table 8.B.6 Reliabilities and SEMs by Disability

Content Area	Level	MR/ID			Hard of Hearing			Deafness			Speech Impairment		
		N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM
English– Language Arts	I	5,572	0.89	3.71	83	0.86	3.94	39	0.85	4.30	112	0.86	2.81
	II	1,870	0.82	2.36	36	0.89	2.25	29	0.89	2.48	429	0.76	2.34
	III	2,430	0.85	2.29	51	0.86	2.28	43	0.88	2.22	358	0.80	2.21
	IV	4,098	0.89	2.28	68	0.91	2.30	84	0.83	2.36	266	0.84	2.49
	V	4,468	0.89	2.16	65	0.87	2.32	96	0.85	2.36	177	0.81	2.12
Mathematics	I	5,561	0.85	4.09	83	0.84	4.08	39	0.85	4.28	112	0.66	3.77
	II	1,866	0.84	2.41	35	0.88	2.52	29	0.65	3.41	428	0.79	2.33
	III	2,427	0.82	2.49	51	0.80	2.47	43	0.84	2.23	358	0.71	2.56
	IV	4,081	0.82	2.59	68	0.84	2.69	83	0.83	2.51	266	0.82	2.44
	V	4,455	0.86	2.62	65	0.87	2.56	95	0.81	2.58	176	0.74	2.75
Science	I	1,475	0.89	3.75	26	0.85	3.60	12	0.87	4.22	9	–	–
	III	1,281	0.81	2.25	24	0.89	2.12	19	0.53	2.35	177	0.65	2.24
	IV	1,388	0.85	2.36	28	0.85	2.74	27	0.87	2.24	67	0.84	2.17
	V	1,491	0.81	2.32	19	0.89	2.23	33	0.71	2.45	68	0.74	2.41
Content Area	Level	Visual Impairment			Emotional Disturbance			Orthopedic Impairment			Other Health Impairment		
		N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM
English– Language Arts	I	220	0.90	3.91	33	0.95	1.80	1,737	0.89	3.97	419	0.91	3.61
	II	24	0.86	2.29	26	0.84	2.18	189	0.86	2.38	320	0.83	2.33
	III	32	0.92	2.35	46	0.88	2.16	222	0.87	2.30	377	0.87	2.12
	IV	54	0.90	2.48	63	0.83	2.43	444	0.91	2.34	445	0.88	2.37
	V	57	0.91	2.30	145	0.79	2.29	458	0.92	2.16	461	0.88	2.13
Mathematics	I	218	0.90	3.80	32	0.90	2.78	1,726	0.88	3.99	417	0.89	3.92
	II	23	0.87	2.41	26	0.81	2.08	189	0.87	2.38	321	0.84	2.41
	III	32	0.87	2.73	46	0.88	2.34	220	0.83	2.55	375	0.82	2.56
	IV	54	0.86	2.67	62	0.81	2.75	442	0.86	2.61	443	0.83	2.64
	V	57	0.90	2.55	142	0.71	2.73	456	0.90	2.61	465	0.86	2.61
Science	I	86	0.91	3.83	11	0.96	2.52	478	0.90	3.82	86	0.94	3.33
	III	15	0.76	2.24	22	0.90	1.99	115	0.84	2.37	187	0.83	2.21
	IV	21	0.82	2.48	21	0.82	2.30	166	0.83	2.38	140	0.86	2.25
	V	25	0.89	2.38	45	0.74	2.45	141	0.88	2.36	159	0.78	2.38

Content Area	Level	Specific Learning Disability			Deaf-Blindness			Multiple Disabilities			Autism		
		N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM	N	Reliab.	SEM
English– Language Arts	I	86	0.91	2.79	20	0.84	4.28	1,807	0.89	3.97	4,343	0.82	4.14
	II	377	0.69	2.36	1	–	–	115	0.88	2.34	2,244	0.85	2.38
	III	586	0.80	2.05	9	–	–	147	0.90	2.35	2,494	0.90	2.27
	IV	768	0.75	2.33	4	–	–	320	0.91	2.31	3,270	0.89	2.41
	V	839	0.77	2.09	4	–	–	446	0.92	2.19	2,793	0.91	2.26
Mathematics	I	84	0.88	3.17	19	0.78	4.22	1,796	0.88	3.97	4,329	0.80	4.25
	II	377	0.74	2.27	1	–	–	115	0.85	2.47	2,243	0.85	2.47
	III	586	0.69	2.46	8	–	–	147	0.86	2.56	2,487	0.84	2.59
	IV	768	0.69	2.55	4	–	–	320	0.88	2.52	3,261	0.85	2.66
	V	836	0.73	2.57	4	–	–	443	0.90	2.56	2,778	0.87	2.67
Science	I	20	0.60	1.67	3	–	–	502	0.91	3.76	1,003	0.84	4.04
	III	309	0.71	2.09	6	–	–	78	0.85	2.41	1,233	0.85	2.30
	IV	259	0.73	2.12	1	–	–	118	0.89	2.38	1,000	0.87	2.40
	V	293	0.66	2.38	0	–	–	153	0.87	2.36	915	0.88	2.31
Content Area	Level	Traumatic Brain Injury			Unknown Disability								
		N	Reliab.	SEM	N	Reliab.	SEM						
English– Language Arts	I	103	0.93	3.66	348	0.88	3.91						
	II	32	0.84	2.38	180	0.86	2.42						
	III	41	0.90	2.24	132	0.87	2.38						
	IV	67	0.85	2.46	183	0.89	2.46						
	V	66	0.87	2.14	293	0.91	2.19						
Mathematics	I	102	0.92	3.81	348	0.86	4.12						
	II	32	0.83	2.24	179	0.88	2.40						
	III	41	0.87	2.43	131	0.83	2.67						
	IV	68	0.80	2.79	183	0.85	2.67						
	V	66	0.89	2.46	286	0.86	2.70						
Science	I	30	0.92	3.78	59	0.88	3.91						
	III	15	0.73	2.40	70	0.87	2.31						
	IV	20	0.65	2.38	34	0.90	2.34						
	V	26	0.85	2.21	82	0.88	2.27						

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

	Placement Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–3	0.02	0.02	0.00	0.00	0.00	0.05
	4–8	0.00	0.03	0.02	0.01	0.00	0.06
	9–13	0.00	0.01	0.03	0.03	0.00	0.07
<b>All-forms Average *</b>	14–24	0.00	0.01	0.03	0.16	0.06	0.25
	25–40	0.00	0.00	0.00	0.04	0.54	0.58
<b>Estimated Proportion Correctly Classified: Total = 0.77, Proficient &amp; Above = 0.93</b>							
<b>Decision Consistency</b>	0–3	0.02	0.01	0.01	0.00	0.00	0.05
	4–8	0.01	0.02	0.02	0.01	0.00	0.06
	9–13	0.00	0.01	0.02	0.03	0.00	0.07
<b>Alternate Form *</b>	14–24	0.00	0.01	0.04	0.13	0.07	0.25
	25–40	0.00	0.00	0.00	0.06	0.52	0.58
<b>Estimated Proportion Consistently Classified: Total = 0.71, Proficient &amp; Above = 0.91</b>							

\* 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	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–4	0.03	0.03	0.01	0.00	0.00	0.07
	5–10	0.01	0.05	0.03	0.00	0.00	0.09
	11–18	0.00	0.02	0.10	0.04	0.00	0.16
<b>All-forms Average *</b>	19–28	0.00	0.01	0.05	0.19	0.06	0.31
	29–40	0.00	0.00	0.00	0.05	0.33	0.38
<b>Estimated Proportion Correctly Classified: Total = 0.70, Proficient &amp; Above = 0.90</b>							
<b>Decision Consistency</b>	0–4	0.03	0.02	0.01	0.00	0.00	0.07
	5–10	0.02	0.04	0.03	0.00	0.00	0.09
	11–18	0.00	0.03	0.07	0.05	0.00	0.16
<b>Alternate Form *</b>	19–28	0.00	0.01	0.06	0.15	0.08	0.31
	29–40	0.00	0.00	0.00	0.07	0.32	0.38
<b>Estimated Proportion Consistently Classified: Total = 0.61, Proficient &amp; Above = 0.86</b>							

\* 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	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–5	0.05	0.02	0.00	0.00	0.00	0.08
	6–10	0.01	0.04	0.03	0.00	0.00	0.08
	11–19	0.00	0.02	0.11	0.04	0.00	0.17
<b>All-forms Average *</b>	20–29	0.00	0.00	0.04	0.16	0.05	0.25
	30–40	0.00	0.00	0.00	0.05	0.36	0.41
<b>Estimated Proportion Correctly Classified: Total = 0.73, Proficient &amp; Above = 0.91</b>							
<b>Decision Consistency</b>	0–5	0.05	0.02	0.01	0.00	0.00	0.08
	6–10	0.02	0.03	0.03	0.00	0.00	0.08
	11–19	0.01	0.03	0.08	0.05	0.00	0.17
<b>Alternate Form *</b>	20–29	0.00	0.01	0.05	0.13	0.06	0.25
	30–40	0.00	0.00	0.00	0.06	0.35	0.41
<b>Estimated Proportion Consistently Classified: Total = 0.64, Proficient &amp; Above = 0.88</b>							

\* 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	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–2	0.00	0.01	0.00	0.00	0.00	0.01
	3–8	0.00	0.03	0.02	0.00	0.00	0.05
	9–13	0.00	0.01	0.10	0.05	0.00	0.15
<b>All-forms Average *</b>	14–19	0.00	0.01	0.05	0.24	0.07	0.37
	20–32	0.00	0.00	0.00	0.05	0.37	0.42
<b>Estimated Proportion Correctly Classified: Total = 0.74, Proficient &amp; Above = 0.90</b>							
<b>Decision Consistency</b>	0–2	0.00	0.01	0.00	0.00	0.00	0.01
	3–8	0.00	0.03	0.02	0.00	0.00	0.05
	9–13	0.00	0.02	0.08	0.05	0.00	0.15
<b>Alternate Form *</b>	14–19	0.00	0.01	0.06	0.20	0.10	0.37
	20–32	0.00	0.00	0.00	0.07	0.35	0.42
<b>Estimated Proportion Consistently Classified: Total = 0.66, Proficient &amp; Above = 0.87</b>							

\* 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	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–7	0.02	0.01	0.00	0.00	0.00	0.03
	8–13	0.01	0.09	0.04	0.00	0.00	0.14
	14–17	0.00	0.03	0.08	0.06	0.00	0.18
<b>All-forms Average *</b>	18–23	0.00	0.00	0.04	0.23	0.05	0.32
	24–32	0.00	0.00	0.00	0.06	0.27	0.33
<b>Estimated Proportion Correctly Classified: Total = 0.69, Proficient &amp; Above = 0.89</b>							
<b>Decision Consistency</b>	0–7	0.02	0.01	0.00	0.00	0.00	0.03
	8–13	0.02	0.07	0.04	0.01	0.00	0.14
	14–17	0.00	0.04	0.06	0.06	0.01	0.18
<b>Alternate Form *</b>	18–23	0.00	0.01	0.06	0.18	0.08	0.32
	24–32	0.00	0.00	0.00	0.08	0.25	0.33
<b>Estimated Proportion Consistently Classified: Total = 0.59, Proficient &amp; Above = 0.85</b>							

\* 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	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–3	0.00	0.01	0.00	0.00	0.00	0.02
	4–7	0.00	0.02	0.01	0.00	0.00	0.02
	8–12	0.00	0.01	0.07	0.03	0.00	0.11
<b>All-forms Average *</b>	13–20	0.00	0.00	0.03	0.26	0.05	0.34
	21–32	0.00	0.00	0.00	0.05	0.46	0.50
<b>Estimated Proportion Correctly Classified: Total = 0.80, Proficient &amp; Above = 0.94</b>							
<b>Decision Consistency</b>	0–3	0.01	0.01	0.00	0.00	0.00	0.02
	4–7	0.00	0.01	0.01	0.00	0.00	0.02
	8–12	0.00	0.02	0.06	0.04	0.00	0.11
<b>Alternate Form *</b>	13–20	0.00	0.00	0.04	0.22	0.07	0.34
	21–32	0.00	0.00	0.00	0.07	0.43	0.50
<b>Estimated Proportion Consistently Classified: Total = 0.73, Proficient &amp; Above = 0.91</b>							

\* 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	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–3	0.00	0.01	0.00	0.00	0.00	0.01
	4–10	0.00	0.04	0.02	0.00	0.00	0.07
	11–17	0.00	0.02	0.19	0.07	0.01	0.29
<b>All-forms Average *</b>	18–24	0.00	0.00	0.05	0.28	0.04	0.37
	25–32	0.00	0.00	0.00	0.07	0.20	0.27
<b>Estimated Proportion Correctly Classified: Total = 0.71, Proficient &amp; Above = 0.88</b>							
<b>Decision Consistency</b>	0–3	0.00	0.01	0.00	0.00	0.00	0.01
	4–10	0.00	0.04	0.02	0.00	0.00	0.07
	11–17	0.00	0.04	0.15	0.08	0.01	0.29
<b>Alternate Form *</b>	18–24	0.00	0.00	0.07	0.23	0.07	0.37
	25–32	0.00	0.00	0.01	0.08	0.19	0.27
<b>Estimated Proportion Consistently Classified: Total = 0.61, Proficient &amp; Above = 0.83</b>							

\* 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	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–3	0.00	0.00	0.00	0.00	0.00	0.01
	4–10	0.00	0.03	0.01	0.00	0.00	0.04
	11–18	0.00	0.01	0.18	0.05	0.00	0.25
<b>All-forms Average *</b>	19–26	0.00	0.00	0.05	0.45	0.05	0.54
	27–32	0.00	0.00	0.00	0.07	0.09	0.16
<b>Estimated Proportion Correctly Classified: Total = 0.75, Proficient &amp; Above = 0.90</b>							
<b>Decision Consistency</b>	0–3	0.00	0.00	0.00	0.00	0.00	0.01
	4–10	0.00	0.03	0.01	0.00	0.00	0.04
	11–18	0.00	0.02	0.16	0.07	0.00	0.25
<b>Alternate Form *</b>	19–26	0.00	0.00	0.07	0.38	0.08	0.54
	27–32	0.00	0.00	0.00	0.08	0.08	0.16
<b>Estimated Proportion Consistently Classified: Total = 0.66, Proficient &amp; Above = 0.86</b>							

\* 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	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–4	0.01	0.02	0.00	0.00	0.00	0.03
	5–8	0.00	0.04	0.02	0.00	0.00	0.07
	9–12	0.00	0.02	0.08	0.05	0.01	0.16
<b>All-forms Average *</b>	13–20	0.00	0.00	0.03	0.26	0.04	0.33
	21–32	0.00	0.00	0.00	0.05	0.36	0.41
<b>Estimated Proportion Correctly Classified: Total = 0.75, Proficient &amp; Above = 0.91</b>							
<b>Decision Consistency</b>	0–4	0.01	0.01	0.00	0.00	0.00	0.03
	5–8	0.01	0.03	0.02	0.00	0.00	0.07
	9–12	0.01	0.03	0.06	0.06	0.01	0.16
<b>Alternate Form *</b>	13–20	0.00	0.00	0.04	0.22	0.06	0.33
	21–32	0.00	0.00	0.00	0.07	0.34	0.41
<b>Estimated Proportion Consistently Classified: Total = 0.67, Proficient &amp; Above = 0.88</b>							

\* 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	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–5	0.01	0.01	0.00	0.00	0.00	0.02
	6–10	0.01	0.06	0.04	0.00	0.00	0.11
	11–15	0.00	0.03	0.12	0.07	0.00	0.22
<b>All-forms Average *</b>	16–22	0.00	0.00	0.04	0.26	0.04	0.35
	23–32	0.00	0.00	0.00	0.07	0.23	0.30
<b>Estimated Proportion Correctly Classified: Total = 0.69, Proficient &amp; Above = 0.88</b>							
<b>Decision Consistency</b>	0–5	0.01	0.01	0.00	0.00	0.00	0.02
	6–10	0.01	0.05	0.04	0.01	0.00	0.11
	11–15	0.00	0.04	0.09	0.07	0.01	0.22
<b>Alternate Form *</b>	16–22	0.00	0.00	0.06	0.21	0.07	0.35
	23–32	0.00	0.00	0.01	0.08	0.22	0.30
<b>Estimated Proportion Consistently Classified: Total = 0.58, Proficient &amp; Above = 0.84</b>							

\* 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	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–3	0.00	0.01	0.00	0.00	0.00	0.01
	4–12	0.00	0.06	0.02	0.00	0.00	0.09
	13–19	0.00	0.02	0.18	0.05	0.00	0.25
<b>All-forms Average *</b>	20–27	0.00	0.00	0.06	0.38	0.04	0.48
	28–32	0.00	0.00	0.00	0.05	0.13	0.18
<b>Estimated Proportion Correctly Classified: Total = 0.75, Proficient &amp; Above = 0.89</b>							
<b>Decision Consistency</b>	0–3	0.00	0.01	0.00	0.00	0.00	0.01
	4–12	0.00	0.06	0.03	0.00	0.00	0.09
	13–19	0.00	0.03	0.15	0.07	0.00	0.25
<b>Alternate Form *</b>	20–27	0.00	0.01	0.08	0.32	0.07	0.48
	28–32	0.00	0.00	0.00	0.06	0.12	0.18
<b>Estimated Proportion Consistently Classified: Total = 0.66, Proficient &amp; Above = 0.84</b>							

\* 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	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–3	0.01	0.01	0.00	0.00	0.00	0.02
	4–8	0.00	0.03	0.01	0.00	0.00	0.03
	9–14	0.00	0.01	0.10	0.04	0.00	0.16
<b>All-forms Average *</b>	15–22	0.00	0.00	0.03	0.26	0.05	0.35
	23–32	0.00	0.00	0.00	0.05	0.40	0.44
<b>Estimated Proportion Correctly Classified: Total = 0.79, Proficient &amp; Above = 0.93</b>							
<b>Decision Consistency</b>	0–3	0.01	0.01	0.00	0.00	0.00	0.02
	4–8	0.00	0.02	0.01	0.00	0.00	0.03
	9–14	0.00	0.02	0.09	0.05	0.00	0.16
<b>Alternate Form *</b>	15–22	0.00	0.00	0.05	0.23	0.07	0.35
	23–32	0.00	0.00	0.00	0.06	0.38	0.44
<b>Estimated Proportion Consistently Classified: Total = 0.72, Proficient &amp; Above = 0.90</b>							

\* 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	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–5	0.01	0.01	0.00	0.00	0.00	0.02
	6–10	0.01	0.05	0.03	0.00	0.00	0.09
	11–15	0.00	0.02	0.09	0.05	0.00	0.17
<b>All-forms Average *</b>	16–23	0.00	0.00	0.04	0.26	0.04	0.34
	24–32	0.00	0.00	0.00	0.06	0.32	0.37
<b>Estimated Proportion Correctly Classified: Total = 0.73, Proficient &amp; Above = 0.91</b>							
<b>Decision Consistency</b>	0–5	0.01	0.01	0.00	0.00	0.00	0.02
	6–10	0.01	0.04	0.03	0.01	0.00	0.09
	11–15	0.00	0.03	0.07	0.06	0.00	0.17
<b>Alternate Form *</b>	16–23	0.00	0.01	0.05	0.21	0.07	0.34
	24–32	0.00	0.00	0.00	0.07	0.30	0.37
<b>Estimated Proportion Consistently Classified: Total = 0.64, Proficient &amp; Above = 0.87</b>							

\* 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	Far Below Basic	Below Basic	Basic	Proficient	Advanced	Category Total †
<b>Decision Accuracy</b>	0–3	0.00	0.01	0.00	0.00	0.00	0.01
	4–10	0.00	0.04	0.01	0.00	0.00	0.05
	11–18	0.00	0.02	0.21	0.06	0.00	0.30
<b>All-forms Average *</b>	19–24	0.00	0.00	0.05	0.28	0.07	0.41
	25–32	0.00	0.00	0.00	0.05	0.19	0.24
<b>Estimated Proportion Correctly Classified: Total = 0.72, Proficient &amp; Above = 0.88</b>							
<b>Decision Consistency</b>	0–3	0.00	0.01	0.00	0.00	0.00	0.01
	4–10	0.00	0.03	0.01	0.00	0.00	0.05
	11–18	0.00	0.03	0.18	0.08	0.01	0.29
<b>Alternate Form *</b>	19–24	0.00	0.00	0.08	0.23	0.10	0.41
	25–32	0.00	0.00	0.00	0.07	0.17	0.24
<b>Estimated Proportion Consistently Classified: Total = 0.62, Proficient &amp; Above = 0.84</b>							

\* Values in table are proportions of the total sample.

† Inconsistencies with category cell entries are due to rounding.

## Appendix 8.C—Validity Analyses

Note that, while the correlations are reported only for samples that comprise 11 or more examinees, results based on samples that contain 50 or fewer examinees should be interpreted with caution due to small sample sizes. Correlations between scores on any two content-area tests where 10 or fewer examinees took the tests are expressed as hyphens. Correlations between scores on two content-area tests that cannot be administered to the same group of students are expressed as “N/A.”

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

	Male			Female			Unknown		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>9,470</b>	0.79	0.79	<b>5,357</b>	0.82	0.82	<b>95</b>	0.86	0.89
<b>Mathematics</b>	9,431	<b>9,435</b>	0.79	5,335	<b>5,335</b>	0.81	94	<b>96</b>	0.84
<b>Science</b>	2,347	2,345	<b>2,352</b>	1,430	1,430	<b>1,431</b>	17	17	<b>17</b>

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

	Male			Female			Unknown		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>4,086</b>	0.71	N/A	<b>1,729</b>	0.71	N/A	<b>57</b>	0.74	N/A
<b>Mathematics</b>	4,079	<b>4,084</b>	N/A	1,719	<b>1,723</b>	N/A	57	<b>57</b>	N/A
<b>Science</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>

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

	Male			Female			Unknown		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>4,743</b>	0.76	0.75	<b>2,191</b>	0.74	0.71	<b>34</b>	0.68	0.54
<b>Mathematics</b>	4,729	<b>4,732</b>	0.72	2,188	<b>2,188</b>	0.72	32	<b>32</b>	0.59
<b>Science</b>	2,414	2,413	<b>2,422</b>	1,109	1,108	<b>1,114</b>	15	15	<b>15</b>

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

	Male			Female			Unknown		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>6,670</b>	0.76	0.71	<b>3,406</b>	0.75	0.70	<b>58</b>	0.81	–
<b>Mathematics</b>	6,645	<b>6,650</b>	0.70	3,392	<b>3,395</b>	0.70	58	<b>58</b>	–
<b>Science</b>	2,121	2,121	<b>2,130</b>	1,144	1,141	<b>1,153</b>	7	7	<b>7</b>

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

	Male			Female			Unknown		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>6,799</b>	0.76	0.73	<b>3,518</b>	0.78	0.75	<b>51</b>	0.91	0.91
<b>Mathematics</b>	6,754	<b>6,770</b>	0.71	3,502	<b>3,504</b>	0.71	50	<b>50</b>	0.85
<b>Science</b>	2,279	2,279	<b>2,291</b>	1,142	1,143	<b>1,148</b>	11	11	<b>11</b>

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
<b>ELA</b>	<b>108</b>	0.77	0.83	<b>1,284</b>	0.79	0.78	<b>79</b>	0.82	0.80	<b>429</b>	0.77	0.71
<b>Mathematics</b>	106	<b>106</b>	0.83	1,279	<b>1,279</b>	0.75	78	<b>78</b>	0.86	429	<b>429</b>	0.81
<b>Science</b>	27	27	<b>27</b>	306	306	<b>306</b>	18	18	<b>18</b>	111	111	<b>111</b>
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>7,983</b>	0.81	0.81	<b>1,252</b>	0.82	0.82	<b>3,208</b>	0.80	0.80	<b>579</b>	0.78	0.84
<b>Mathematics</b>	7,954	<b>7,956</b>	0.81	1,242	<b>1,242</b>	0.82	3,197	<b>3,198</b>	0.76	575	<b>578</b>	0.77
<b>Science</b>	2,042	2,042	<b>2,044</b>	335	333	<b>335</b>	835	835	<b>838</b>	120	120	<b>121</b>

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
<b>ELA</b>	<b>40</b>	0.71	N/A	<b>424</b>	0.67	N/A	<b>28</b>	0.69	N/A	<b>177</b>	0.72	N/A
<b>Mathematics</b>	40	<b>40</b>	N/A	424	<b>424</b>	N/A	27	<b>27</b>	N/A	177	<b>179</b>	N/A
<b>Science</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>3,263</b>	0.72	N/A	<b>470</b>	0.72	N/A	<b>1,230</b>	0.72	N/A	<b>240</b>	0.72	N/A
<b>Mathematics</b>	3,254	<b>3,256</b>	N/A	470	<b>473</b>	N/A	1,223	<b>1,225</b>	N/A	240	<b>240</b>	N/A
<b>Science</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>

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
<b>ELA</b>	<b>47</b>	0.75	0.66	<b>483</b>	0.75	0.76	<b>25</b>	0.72	0.65	<b>224</b>	0.71	0.74
<b>Mathematics</b>	47	<b>47</b>	0.59	482	<b>482</b>	0.80	24	<b>24</b>	0.46	224	<b>224</b>	0.72
<b>Science</b>	23	23	<b>23</b>	252	252	<b>252</b>	13	13	<b>13</b>	115	115	<b>116</b>
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>3,867</b>	0.76	0.74	<b>572</b>	0.80	0.74	<b>1,521</b>	0.75	0.72	<b>229</b>	0.74	0.66
<b>Mathematics</b>	3,860	<b>3,861</b>	0.71	569	<b>569</b>	0.76	1,515	<b>1,516</b>	0.71	228	<b>229</b>	0.71
<b>Science</b>	1,931	1,929	<b>1,935</b>	294	294	<b>296</b>	787	787	<b>790</b>	123	123	<b>126</b>

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
<b>ELA</b>	<b>96</b>	0.80	0.72	<b>719</b>	0.75	0.71	<b>57</b>	0.78	0.82	<b>296</b>	0.79	0.72
<b>Mathematics</b>	96	<b>96</b>	0.75	713	<b>714</b>	0.72	57	<b>57</b>	0.87	295	<b>295</b>	0.80
<b>Science</b>	31	31	<b>31</b>	226	226	<b>226</b>	20	20	<b>20</b>	108	108	<b>109</b>
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>5,385</b>	0.76	0.67	<b>932</b>	0.76	0.79	<b>2,237</b>	0.77	0.72	<b>412</b>	0.79	0.79
<b>Mathematics</b>	5,369	<b>5,371</b>	0.68	928	<b>929</b>	0.74	2,227	<b>2,230</b>	0.70	410	<b>411</b>	0.79
<b>Science</b>	1,723	1,722	<b>1,732</b>	294	292	<b>297</b>	771	771	<b>775</b>	99	99	<b>100</b>

**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
<b>ELA</b>	<b>98</b>	0.78	0.60	<b>741</b>	0.76	0.76	<b>57</b>	0.74	0.78	<b>343</b>	0.77	0.69
<b>Mathematics</b>	97	<b>97</b>	0.59	739	<b>740</b>	0.74	57	<b>57</b>	0.90	340	<b>340</b>	0.70
<b>Science</b>	38	38	<b>38</b>	255	255	<b>257</b>	19	19	<b>19</b>	115	115	<b>116</b>
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>5,145</b>	0.76	0.76	<b>1,094</b>	0.75	0.74	<b>2,552</b>	0.75	0.70	<b>338</b>	0.81	0.79
<b>Mathematics</b>	5,118	<b>5,121</b>	0.70	1,080	<b>1,082</b>	0.72	2,541	<b>2,550</b>	0.70	334	<b>337</b>	0.80
<b>Science</b>	1,691	1,690	<b>1,697</b>	363	362	<b>365</b>	828	830	<b>834</b>	123	124	<b>124</b>

**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
<b>ELA</b>	<b>67</b>	0.78	0.73	<b>476</b>	0.81	0.72	<b>41</b>	0.73	–	<b>155</b>	0.71	0.62
<b>Mathematics</b>	66	<b>66</b>	0.94	473	<b>473</b>	0.70	40	<b>40</b>	–	155	<b>155</b>	0.81
<b>Science</b>	16	16	<b>16</b>	125	125	<b>125</b>	9	9	<b>9</b>	41	41	<b>41</b>
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>5,972</b>	0.80	0.81	<b>776</b>	0.80	0.79	<b>1,104</b>	0.80	0.83	<b>237</b>	0.78	0.79
<b>Mathematics</b>	5,952	<b>5,954</b>	0.81	770	<b>770</b>	0.78	1,103	<b>1,104</b>	0.75	235	<b>236</b>	0.72
<b>Science</b>	1,557	1,557	<b>1,559</b>	213	213	<b>213</b>	260	260	<b>260</b>	50	50	<b>51</b>

**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
<b>ELA</b>	<b>23</b>	0.61	N/A	<b>181</b>	0.67	N/A	<b>12</b>	0.53	N/A	<b>58</b>	0.67	N/A
<b>Mathematics</b>	23	<b>23</b>	N/A	181	<b>181</b>	N/A	11	<b>11</b>	N/A	58	<b>59</b>	N/A
<b>Science</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>2,485</b>	0.72	N/A	<b>314</b>	0.74	N/A	<b>517</b>	0.70	N/A	<b>99</b>	0.71	N/A
<b>Mathematics</b>	2,477	<b>2,479</b>	N/A	314	<b>316</b>	N/A	513	<b>514</b>	N/A	99	<b>99</b>	N/A
<b>Science</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>

**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
<b>ELA</b>	<b>32</b>	0.78	0.68	<b>216</b>	0.77	0.73	<b>15</b>	0.80	–	<b>71</b>	0.62	0.82
<b>Mathematics</b>	32	<b>32</b>	0.46	215	<b>215</b>	0.75	14	<b>14</b>	–	71	<b>71</b>	0.71
<b>Science</b>	13	13	<b>13</b>	108	108	<b>108</b>	9	9	<b>9</b>	34	34	<b>34</b>
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>3,077</b>	0.75	0.74	<b>390</b>	0.79	0.72	<b>646</b>	0.73	0.64	<b>101</b>	0.74	0.63
<b>Mathematics</b>	3,072	<b>3,073</b>	0.70	390	<b>390</b>	0.72	645	<b>646</b>	0.65	101	<b>101</b>	0.81
<b>Science</b>	1,563	1,561	<b>1,565</b>	209	209	<b>210</b>	334	334	<b>335</b>	55	55	<b>55</b>

**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
<b>ELA</b>	<b>61</b>	0.76	0.44	<b>299</b>	0.72	0.78	<b>35</b>	0.76	0.81	<b>113</b>	0.78	0.83
<b>Mathematics</b>	61	<b>61</b>	0.52	298	<b>299</b>	0.73	35	<b>35</b>	0.94	112	<b>112</b>	0.84
<b>Science</b>	17	17	<b>17</b>	97	97	<b>97</b>	14	14	<b>14</b>	33	33	<b>33</b>
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>4,288</b>	0.75	0.66	<b>624</b>	0.73	0.77	<b>907</b>	0.75	0.73	<b>186</b>	0.81	0.84
<b>Mathematics</b>	4,274	<b>4,274</b>	0.67	623	<b>624</b>	0.69	903	<b>904</b>	0.70	185	<b>186</b>	0.77
<b>Science</b>	1,386	1,386	<b>1,393</b>	198	198	<b>200</b>	323	323	<b>324</b>	53	53	<b>53</b>

**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
<b>ELA</b>	<b>66</b>	0.66	0.23	<b>309</b>	0.78	0.81	<b>35</b>	0.75	0.83	<b>122</b>	0.80	0.79
<b>Mathematics</b>	65	<b>65</b>	0.38	309	<b>309</b>	0.79	35	<b>35</b>	0.92	121	<b>121</b>	0.78
<b>Science</b>	25	25	<b>25</b>	100	100	<b>101</b>	15	15	<b>15</b>	40	40	<b>41</b>
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>4,001</b>	0.76	0.76	<b>749</b>	0.76	0.74	<b>973</b>	0.74	0.66	<b>122</b>	0.81	0.64
<b>Mathematics</b>	3,979	<b>3,982</b>	0.71	740	<b>741</b>	0.72	970	<b>973</b>	0.66	122	<b>124</b>	0.75
<b>Science</b>	1,319	1,319	<b>1,322</b>	243	243	<b>245</b>	318	319	<b>321</b>	40	40	<b>40</b>

**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
<b>ELA</b>	<b>34</b>	0.71	0.93	<b>734</b>	0.77	0.83	<b>31</b>	0.89	–	<b>257</b>	0.80	0.72
<b>Mathematics</b>	34	<b>34</b>	0.77	732	<b>732</b>	0.77	31	<b>31</b>	–	257	<b>257</b>	0.83
<b>Science</b>	11	11	<b>11</b>	168	168	<b>168</b>	8	8	<b>8</b>	66	66	<b>66</b>
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>1,698</b>	0.84	0.81	<b>398</b>	0.85	0.85	<b>1,947</b>	0.80	0.78	<b>213</b>	0.78	0.85
<b>Mathematics</b>	1,692	<b>1,692</b>	0.83	395	<b>395</b>	0.85	1,938	<b>1,938</b>	0.76	212	<b>212</b>	0.82
<b>Science</b>	431	431	<b>431</b>	104	102	<b>104</b>	548	548	<b>550</b>	49	49	<b>49</b>

**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
<b>ELA</b>	<b>11</b>	0.83	N/A	<b>207</b>	0.66	N/A	<b>14</b>	0.73	N/A	<b>111</b>	0.74	N/A
<b>Mathematics</b>	11	<b>11</b>	N/A	207	<b>207</b>	N/A	14	<b>14</b>	N/A	111	<b>112</b>	N/A
<b>Science</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>569</b>	0.72	N/A	<b>110</b>	0.65	N/A	<b>623</b>	0.71	N/A	<b>77</b>	0.78	N/A
<b>Mathematics</b>	568	<b>568</b>	N/A	110	<b>110</b>	N/A	622	<b>623</b>	N/A	77	<b>77</b>	N/A
<b>Science</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>

**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
<b>ELA</b>	<b>13</b>	0.85	–	<b>247</b>	0.75	0.78	<b>8</b>	–	–	<b>141</b>	0.74	0.70
<b>Mathematics</b>	13	<b>13</b>	–	247	<b>247</b>	0.84	8	<b>8</b>	–	141	<b>141</b>	0.72
<b>Science</b>	9	9	<b>9</b>	136	136	<b>136</b>	2	2	<b>2</b>	79	79	<b>80</b>

	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>651</b>	0.76	0.70	<b>153</b>	0.79	0.81	<b>809</b>	0.76	0.77	<b>79</b>	0.81	0.73
<b>Mathematics</b>	649	<b>649</b>	0.71	152	<b>152</b>	0.83	805	<b>805</b>	0.76	78	<b>79</b>	0.63
<b>Science</b>	314	314	<b>316</b>	76	76	<b>77</b>	425	425	<b>425</b>	44	44	<b>47</b>

**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
<b>ELA</b>	<b>33</b>	0.87	0.87	<b>387</b>	0.76	0.64	<b>17</b>	0.83	–	<b>174</b>	0.78	0.63
<b>Mathematics</b>	33	<b>33</b>	0.88	384	<b>384</b>	0.72	17	<b>17</b>	–	174	<b>174</b>	0.75
<b>Science</b>	13	13	<b>13</b>	128	128	<b>128</b>	5	5	<b>5</b>	71	71	<b>72</b>

	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>916</b>	0.78	0.71	<b>245</b>	0.81	0.81	<b>1,214</b>	0.78	0.73	<b>122</b>	0.78	0.72
<b>Mathematics</b>	914	<b>915</b>	0.70	244	<b>244</b>	0.80	1,210	<b>1,212</b>	0.71	122	<b>122</b>	0.84
<b>Science</b>	303	302	<b>304</b>	80	80	<b>81</b>	419	419	<b>422</b>	27	27	<b>27</b>

**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
<b>ELA</b>	<b>31</b>	0.84	0.90	<b>403</b>	0.74	0.72	<b>20</b>	0.77	–	<b>209</b>	0.75	0.65
<b>Mathematics</b>	31	<b>31</b>	0.81	401	<b>402</b>	0.71	20	<b>20</b>	–	207	<b>207</b>	0.65
<b>Science</b>	13	13	<b>13</b>	150	150	<b>151</b>	4	4	<b>4</b>	74	74	<b>74</b>

	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>914</b>	0.77	0.71	<b>290</b>	0.71	0.75	<b>1,424</b>	0.77	0.71	<b>119</b>	0.81	0.83
<b>Mathematics</b>	912	<b>912</b>	0.65	288	<b>288</b>	0.76	1,419	<b>1,424</b>	0.72	117	<b>118</b>	0.82
<b>Science</b>	303	302	<b>306</b>	109	109	<b>109</b>	465	466	<b>467</b>	54	55	<b>55</b>

**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
<b>ELA</b>	<b>7</b>	–	–	<b>74</b>	0.81	0.96	<b>7</b>	–	–	<b>17</b>	0.95	–
<b>Mathematics</b>	6	<b>6</b>	–	74	<b>74</b>	0.98	7	<b>7</b>	–	17	<b>17</b>	–
<b>Science</b>	0	0	<b>0</b>	13	13	<b>13</b>	1	1	<b>1</b>	4	4	<b>4</b>

	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>313</b>	0.81	0.81	<b>78</b>	0.84	0.63	<b>157</b>	0.83	0.90	<b>129</b>	0.80	0.92
<b>Mathematics</b>	310	<b>310</b>	0.87	77	<b>77</b>	0.91	156	<b>156</b>	0.91	128	<b>130</b>	0.81
<b>Science</b>	54	54	<b>54</b>	18	18	<b>18</b>	27	27	<b>28</b>	21	21	<b>21</b>

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	6	–	N/A	36	0.77	N/A	2	–	N/A	8	–	N/A
Mathematics	6	6	N/A	36	36	N/A	2	2	N/A	8	8	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	209	0.70	N/A	46	0.77	N/A	90	0.72	N/A	64	0.73	N/A
Mathematics	209	209	N/A	46	47	N/A	88	88	N/A	64	64	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	2	–	–	20	0.84	–	2	–	–	12	0.71	–
Mathematics	2	2	–	20	20	–	2	2	–	12	12	–
Science	1	1	1	8	8	8	2	2	2	2	2	2
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
ELA	139	0.80	0.87	29	0.85	–	66	0.66	0.73	49	0.54	0.67
Mathematics	139	139	0.85	27	27	–	65	65	0.50	49	49	0.46
Science	54	54	54	9	9	9	28	28	30	24	24	24

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	–	–	33	0.81	–	5	–	–	9	–	–
Mathematics	2	2	–	31	31	–	5	5	–	9	9	–
Science	1	1	1	1	1	1	1	1	1	4	4	4
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
ELA	181	0.73	0.63	63	0.78	0.73	116	0.80	0.64	104	0.73	0.84
Mathematics	181	182	0.69	61	61	0.81	114	114	0.66	103	103	0.79
Science	34	34	35	16	14	16	29	29	29	19	19	20

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	1	–	–	29	0.86	–	2	–	–	12	0.02	–
Mathematics	1	1	–	29	29	–	2	2	–	12	12	–
Science	0	0	0	5	5	5	0	0	0	1	1	1
	Hispanic			African American			White			Unknown Ethnicity		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
ELA	230	0.80	0.79	55	0.73	0.60	155	0.65	0.78	97	0.79	0.84
Mathematics	227	227	0.74	52	53	–	152	153	0.67	95	95	0.83
Science	69	69	69	11	10	11	45	45	46	29	29	29

**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
<b>ELA</b>	<b>8,828</b>	0.80	0.80	<b>5,312</b>	0.81	0.80	<b>782</b>	0.82	0.86
<b>Mathematics</b>	8,794	<b>8,798</b>	0.79	5,291	<b>5,291</b>	0.80	775	<b>777</b>	0.89
<b>Science</b>	2,271	2,271	<b>2,274</b>	1,385	1,383	<b>1,387</b>	138	138	<b>139</b>

**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
<b>ELA</b>	<b>3,689</b>	0.71	N/A	<b>1,722</b>	0.71	N/A	<b>461</b>	0.72	N/A
<b>Mathematics</b>	3,676	<b>3,682</b>	N/A	1,720	<b>1,722</b>	N/A	459	<b>460</b>	N/A
<b>Science</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>

**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
<b>ELA</b>	<b>4,548</b>	0.75	0.72	<b>2,101</b>	0.76	0.75	<b>319</b>	0.73	0.76
<b>Mathematics</b>	4,540	<b>4,542</b>	0.70	2,093	<b>2,094</b>	0.75	316	<b>316</b>	0.69
<b>Science</b>	2,325	2,323	<b>2,329</b>	1,085	1,085	<b>1,092</b>	128	128	<b>130</b>

**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
<b>ELA</b>	<b>6,513</b>	0.75	0.70	<b>3,108</b>	0.78	0.72	<b>513</b>	0.76	0.69
<b>Mathematics</b>	6,491	<b>6,495</b>	0.69	3,098	<b>3,101</b>	0.72	506	<b>507</b>	0.73
<b>Science</b>	2,121	2,121	<b>2,131</b>	1,046	1,045	<b>1,052</b>	105	103	<b>107</b>

**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
<b>ELA</b>	<b>6,377</b>	0.76	0.74	<b>3,410</b>	0.76	0.72	<b>581</b>	0.76	0.78
<b>Mathematics</b>	6,341	<b>6,350</b>	0.71	3,395	<b>3,402</b>	0.70	570	<b>572</b>	0.72
<b>Science</b>	2,100	2,101	<b>2,110</b>	1,172	1,173	<b>1,179</b>	160	159	<b>161</b>

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

	MR/ID			Hard of Hearing			Deafness			Speech Impairment		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>5,572</b>	0.77	0.77	<b>83</b>	0.74	0.63	<b>39</b>	0.82	0.87	<b>112</b>	0.60	–
<b>Mathematics</b>	5,559	<b>5,561</b>	0.76	83	<b>83</b>	0.62	39	<b>39</b>	0.86	112	<b>112</b>	–
<b>Science</b>	1,473	1,473	<b>1,475</b>	26	26	<b>26</b>	12	12	<b>12</b>	9	9	<b>9</b>
	Visual Impairment			Emotional Disturbance			Orthopedic Impairment			Other Health Impairment		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>220</b>	0.87	0.89	<b>33</b>	0.75	0.83	<b>1,737</b>	0.85	0.83	<b>419</b>	0.85	0.86
<b>Mathematics</b>	218	<b>218</b>	0.83	32	<b>32</b>	0.86	1,726	<b>1,726</b>	0.84	417	<b>417</b>	0.86
<b>Science</b>	86	86	<b>86</b>	11	11	<b>11</b>	477	476	<b>478</b>	86	86	<b>86</b>
	Specific Learning Disability			Deaf-Blindness			Multiple Disabilities			Autism		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>86</b>	0.80	0.59	<b>20</b>	0.90	–	<b>1,807</b>	0.84	0.85	<b>4,343</b>	0.74	0.71
<b>Mathematics</b>	84	<b>84</b>	0.64	19	<b>19</b>	–	1,796	<b>1,796</b>	0.85	4,327	<b>4,329</b>	0.69
<b>Science</b>	20	20	<b>20</b>	3	3	<b>3</b>	501	501	<b>502</b>	1,001	1,000	<b>1,003</b>
	Traumatic Brain Injury			Unknown Disability								
	ELA	Math	Science	ELA	Math	Science						
<b>ELA</b>	<b>103</b>	0.88	0.86	<b>348</b>	0.82	0.81						
<b>Mathematics</b>	102	<b>102</b>	0.84	346	<b>348</b>	0.79						
<b>Science</b>	30	30	<b>30</b>	59	59	<b>59</b>						

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

	MR/ID			Hard of Hearing			Deafness			Speech Impairment		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>1,870</b>	0.74	N/A	<b>36</b>	0.84	N/A	<b>29</b>	0.67	N/A	<b>429</b>	0.59	N/A
<b>Mathematics</b>	1,861	<b>1,866</b>	N/A	35	<b>35</b>	N/A	29	<b>29</b>	N/A	428	<b>428</b>	N/A
<b>Science</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>
	Visual Impairment			Emotional Disturbance			Orthopedic Impairment			Other Health Impairment		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>24</b>	0.76	N/A	<b>26</b>	0.85	N/A	<b>189</b>	0.75	N/A	<b>320</b>	0.73	N/A
<b>Mathematics</b>	23	<b>23</b>	N/A	26	<b>26</b>	N/A	189	<b>189</b>	N/A	320	<b>321</b>	N/A
<b>Science</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>
	Specific Learning Disability			Deaf-Blindness			Multiple Disabilities			Autism		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>377</b>	0.55	N/A	<b>1</b>	–	N/A	<b>115</b>	0.81	N/A	<b>2,244</b>	0.71	N/A
<b>Mathematics</b>	376	<b>377</b>	N/A	1	<b>1</b>	N/A	115	<b>115</b>	N/A	2,242	<b>2,243</b>	N/A
<b>Science</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>
	Traumatic Brain Injury			Unknown Disability								
	ELA	Math	Science	ELA	Math	Science						
<b>ELA</b>	<b>32</b>	0.47	N/A	<b>180</b>	0.79	N/A						
<b>Mathematics</b>	32	<b>32</b>	N/A	178	<b>179</b>	N/A						
<b>Science</b>	N/A	N/A	<b>N/A</b>	N/A	N/A	<b>N/A</b>						

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

	MR/ID			Hard of Hearing			Deafness			Speech Impairment		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>2,430</b>	0.73	0.70	<b>51</b>	0.76	0.75	<b>43</b>	0.66	0.44	<b>358</b>	0.60	0.51
<b>Mathematics</b>	2,427	<b>2,427</b>	0.73	51	<b>51</b>	0.86	43	<b>43</b>	0.65	358	<b>358</b>	0.51
<b>Science</b>	1,275	1,275	<b>1,281</b>	24	24	<b>24</b>	19	19	<b>19</b>	177	177	<b>177</b>
	Visual Impairment			Emotional Disturbance			Orthopedic Impairment			Other Health Impairment		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>32</b>	0.74	0.87	<b>46</b>	0.78	0.64	<b>222</b>	0.72	0.73	<b>377</b>	0.71	0.65
<b>Mathematics</b>	32	<b>32</b>	0.79	46	<b>46</b>	0.48	220	<b>220</b>	0.77	375	<b>375</b>	0.63
<b>Science</b>	15	15	<b>15</b>	22	22	<b>22</b>	115	113	<b>115</b>	185	185	<b>187</b>
	Specific Learning Disability			Deaf-Blindness			Multiple Disabilities			Autism		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>586</b>	0.57	0.57	<b>9</b>	–	–	<b>147</b>	0.80	0.80	<b>2,494</b>	0.79	0.77
<b>Mathematics</b>	586	<b>586</b>	0.49	8	<b>8</b>	–	146	<b>147</b>	0.77	2,485	<b>2,487</b>	0.75
<b>Science</b>	308	308	<b>309</b>	6	6	<b>6</b>	77	77	<b>78</b>	1,231	1,231	<b>1,233</b>
	Traumatic Brain Injury			Unknown Disability								
	ELA	Math	Science	ELA	Math	Science						
<b>ELA</b>	<b>41</b>	0.77	0.44	<b>132</b>	0.75	0.83						
<b>Mathematics</b>	41	<b>41</b>	0.45	131	<b>131</b>	0.80						
<b>Science</b>	15	15	<b>15</b>	69	69	<b>70</b>						

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

	MR/ID			Hard of Hearing			Deafness			Speech Impairment		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>4,098</b>	0.75	0.70	<b>68</b>	0.73	0.67	<b>84</b>	0.72	0.52	<b>266</b>	0.68	0.62
<b>Mathematics</b>	4,080	<b>4,081</b>	0.71	68	<b>68</b>	0.60	83	<b>83</b>	0.72	266	<b>266</b>	0.61
<b>Science</b>	1,379	1,377	<b>1,388</b>	28	28	<b>28</b>	27	27	<b>27</b>	66	66	<b>67</b>
	Visual Impairment			Emotional Disturbance			Orthopedic Impairment			Other Health Impairment		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>54</b>	0.78	0.80	<b>63</b>	0.68	0.61	<b>444</b>	0.79	0.71	<b>445</b>	0.75	0.69
<b>Mathematics</b>	54	<b>54</b>	0.51	62	<b>62</b>	0.54	442	<b>442</b>	0.70	443	<b>443</b>	0.71
<b>Science</b>	21	21	<b>21</b>	21	21	<b>21</b>	166	166	<b>166</b>	138	138	<b>140</b>
	Specific Learning Disability			Deaf-Blindness			Multiple Disabilities			Autism		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>768</b>	0.55	0.37	<b>4</b>	–	–	<b>320</b>	0.78	0.76	<b>3,270</b>	0.77	0.72
<b>Mathematics</b>	767	<b>768</b>	0.36	4	<b>4</b>	–	319	<b>320</b>	0.81	3,258	<b>3,261</b>	0.72
<b>Science</b>	258	258	<b>259</b>	1	1	<b>1</b>	118	118	<b>118</b>	997	996	<b>1,000</b>
	Traumatic Brain Injury			Unknown Disability								
	ELA	Math	Science	ELA	Math	Science						
<b>ELA</b>	<b>67</b>	0.68	0.56	<b>183</b>	0.79	0.69						
<b>Mathematics</b>	67	<b>68</b>	0.27	182	<b>183</b>	0.76						
<b>Science</b>	20	20	<b>20</b>	32	32	<b>34</b>						

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

	MR/ID			Hard of Hearing			Deafness			Speech Impairment		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>4,468</b>	0.76	0.72	<b>65</b>	0.80	0.82	<b>96</b>	0.63	0.69	<b>177</b>	0.54	0.57
<b>Mathematics</b>	4,451	<b>4,455</b>	0.68	65	<b>65</b>	0.91	95	<b>95</b>	0.62	176	<b>176</b>	0.41
<b>Science</b>	1,484	1,484	<b>1,491</b>	19	19	<b>19</b>	33	33	<b>33</b>	68	67	<b>68</b>
	Visual Impairment			Emotional Disturbance			Orthopedic Impairment			Other Health Impairment		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>57</b>	0.81	0.87	<b>145</b>	0.48	0.35	<b>458</b>	0.83	0.85	<b>461</b>	0.73	0.63
<b>Mathematics</b>	57	<b>57</b>	0.86	142	<b>142</b>	0.34	455	<b>456</b>	0.85	461	<b>465</b>	0.62
<b>Science</b>	25	25	<b>25</b>	45	45	<b>45</b>	140	141	<b>141</b>	157	158	<b>159</b>
	Specific Learning Disability			Deaf-Blindness			Multiple Disabilities			Autism		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
<b>ELA</b>	<b>839</b>	0.56	0.55	<b>4</b>	–	–	<b>446</b>	0.79	0.74	<b>2,793</b>	0.77	0.75
<b>Mathematics</b>	835	<b>836</b>	0.51	4	<b>4</b>	–	442	<b>443</b>	0.77	2,771	<b>2,778</b>	0.74
<b>Science</b>	290	289	<b>293</b>	0	0	<b>0</b>	151	152	<b>153</b>	914	915	<b>915</b>
	Traumatic Brain Injury			Unknown Disability								
	ELA	Math	Science	ELA	Math	Science						
<b>ELA</b>	<b>66</b>	0.78	0.63	<b>293</b>	0.79	0.80						
<b>Mathematics</b>	66	<b>66</b>	0.59	286	<b>286</b>	0.73						
<b>Science</b>	26	26	<b>26</b>	80	79	<b>82</b>						

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	2,149	3.44	1.93	2,149	3.44	1.93	97.16	1.81	1.02	0.05	0.98
	3	2,149	3.43	1.90	2,149	3.42	1.90	96.60	2.09	1.30	0.06	0.98
	4	2,149	3.82	1.79	2,149	3.81	1.79	96.84	2.00	1.16	0.06	0.98
	6	2,149	3.21	1.94	2,149	3.21	1.94	94.93	2.89	2.19	0.10	0.96
	7	2,149	3.74	1.85	2,149	3.73	1.84	94.46	3.72	1.81	0.10	0.96
	9	2,149	3.27	1.95	2,149	3.26	1.95	96.04	2.47	1.49	0.07	0.97
	10	2,149	3.37	1.91	2,149	3.32	1.93	95.21	3.30	1.49	0.09	0.97
	12	2,149	2.90	1.91	2,149	2.92	1.91	93.76	3.58	2.65	0.13	0.95
Mathematics	1	2,130	3.58	1.86	2,130	3.57	1.86	96.53	2.54	0.94	0.06	0.98
	3	2,130	2.83	1.96	2,130	2.82	1.97	95.35	3.38	1.27	0.07	0.98
	4	2,130	2.75	1.90	2,130	2.74	1.90	95.21	2.96	1.83	0.09	0.96
	6	2,130	2.93	1.95	2,130	2.91	1.94	96.38	2.07	1.55	0.08	0.97
	7	2,130	3.13	1.95	2,130	3.12	1.95	96.06	2.02	1.92	0.08	0.97
	9	2,130	3.51	1.90	2,130	3.50	1.90	96.48	2.35	1.17	0.06	0.98
	10	2,130	2.82	1.96	2,130	2.82	1.96	95.45	2.54	2.02	0.09	0.97
	12	2,130	3.48	1.90	2,130	3.50	1.88	94.23	3.00	2.77	0.13	0.94
Science	1	526	3.29	1.93	526	3.29	1.93	96.39	2.47	1.14	0.07	0.97
	3	526	3.35	1.95	526	3.34	1.95	97.15	1.90	0.95	0.06	0.98
	4	526	3.31	1.95	526	3.32	1.93	95.63	3.04	1.33	0.08	0.97
	6	526	3.13	1.97	526	3.12	1.97	96.96	2.09	0.95	0.06	0.98
	7	526	3.48	1.96	526	3.47	1.95	95.82	2.85	1.33	0.07	0.97
	9	526	2.73	1.96	526	2.75	1.95	96.58	2.09	1.33	0.07	0.97
	10	526	3.15	1.95	526	3.14	1.95	95.25	3.99	0.76	0.07	0.98
	12	526	3.34	1.96	526	3.37	1.95	96.20	2.47	1.33	0.07	0.97

\* 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		
<b>English– Language Arts</b>	1	1,328	1.94	1.04	1,328	1.95	1.04	95.26	3.69	1.05	0.06	0.96
	3	1,328	3.63	0.79	1,328	3.63	0.78	96.61	2.56	0.83	0.05	0.91
	4	1,328	1.98	1.07	1,328	1.98	1.08	95.56	3.84	0.60	0.05	0.97
	6	1,328	2.12	1.11	1,328	2.12	1.11	93.75	5.42	0.83	0.07	0.96
	7	1,328	2.45	1.33	1,328	2.44	1.33	94.88	4.29	0.83	0.06	0.97
	9	1,328	2.33	1.10	1,328	2.33	1.10	95.56	3.84	0.60	0.05	0.97
	10	1,328	1.86	1.00	1,328	1.86	1.00	95.11	4.52	0.38	0.05	0.97
	12	1,328	1.87	0.95	1,328	1.87	0.96	93.98	5.35	0.68	0.07	0.95
<b>Mathematics</b>	1	1,330	2.58	0.87	1,330	2.57	0.88	96.84	2.93	0.23	0.03	0.97
	3	1,330	2.99	1.16	1,330	2.98	1.15	97.74	1.80	0.45	0.03	0.99
	4	1,330	3.25	0.88	1,330	3.26	0.88	97.22	2.63	0.15	0.03	0.98
	6	1,330	2.07	1.40	1,330	2.07	1.40	96.84	2.41	0.75	0.04	0.98
	7	1,330	2.94	1.10	1,330	2.94	1.10	97.22	2.26	0.53	0.03	0.98
	9	1,330	2.37	1.34	1,330	2.38	1.35	97.89	1.73	0.38	0.03	0.99
	10	1,330	1.19	0.74	1,330	1.19	0.75	96.69	2.78	0.53	0.04	0.95
	12	1,330	2.77	1.13	1,330	2.78	1.12	94.51	3.98	1.50	0.08	0.93

\* 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 Content Area	Task	First Rating			Second Rating			% Agreement			MAD *	Corr. †
		N	Mean	SD	N	Mean	SD	Exact	Adjacent	Neither		
English– Language Arts	1	1,479	2.41	1.24	1,479	2.41	1.24	95.00	4.67	0.34	0.05	0.98
	3	1,479	2.68	1.06	1,479	2.68	1.07	94.86	4.80	0.34	0.06	0.97
	4	1,479	2.54	1.17	1,479	2.54	1.17	94.86	4.33	0.81	0.06	0.97
	6	1,479	2.26	0.92	1,479	2.26	0.93	95.40	4.19	0.41	0.05	0.96
	7	1,479	3.35	1.07	1,479	3.34	1.08	97.63	1.96	0.41	0.03	0.98
	9	1,479	2.30	0.87	1,479	2.30	0.87	94.52	4.73	0.74	0.06	0.95
	10	1,479	2.30	1.21	1,479	2.29	1.22	95.00	4.60	0.41	0.06	0.98
	12	1,479	2.46	1.19	1,479	2.49	1.17	95.06	3.38	1.56	0.08	0.94
Mathematics	1	1,475	3.18	1.17	1,475	3.19	1.17	98.44	1.36	0.20	0.02	0.99
	3	1,475	2.41	1.04	1,475	2.42	1.04	96.00	3.46	0.54	0.05	0.96
	4	1,475	1.95	1.06	1,475	1.95	1.06	95.05	4.14	0.81	0.06	0.96
	6	1,475	2.87	1.30	1,475	2.88	1.30	97.76	1.36	0.88	0.04	0.97
	7	1,475	2.44	1.38	1,475	2.43	1.38	97.76	1.90	0.34	0.03	0.99
	9	1,475	2.75	0.85	1,475	2.75	0.86	97.76	1.97	0.27	0.03	0.98
	10	1,475	2.15	1.26	1,475	2.15	1.25	96.47	3.12	0.41	0.04	0.98
	12	1,475	2.33	0.89	1,475	2.33	0.89	95.80	3.80	0.41	0.05	0.95
Science	1	724	2.49	1.04	724	2.49	1.03	95.58	4.14	0.28	0.05	0.98
	3	724	2.56	1.01	724	2.57	1.00	95.03	3.31	1.66	0.07	0.95
	4	724	2.22	1.00	724	2.21	0.99	95.03	4.42	0.55	0.06	0.95
	6	724	2.53	0.99	724	2.52	0.98	94.06	5.39	0.55	0.07	0.96
	7	724	2.82	1.01	724	2.83	1.00	95.30	3.87	0.83	0.06	0.95
	9	724	3.00	1.05	724	3.02	1.04	95.86	3.04	1.10	0.06	0.96
	10	724	2.37	1.00	724	2.37	0.99	95.44	3.04	1.52	0.06	0.94
	12	724	3.23	0.92	724	3.23	0.91	97.38	1.93	0.69	0.04	0.93

\* 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	1,591	2.39	1.17	1,591	2.40	1.17	97.30	2.51	0.19	0.03	0.98
	3	1,591	2.48	1.18	1,591	2.49	1.18	94.41	5.15	0.44	0.06	0.97
	4	1,591	2.55	1.00	1,591	2.56	1.00	95.35	4.09	0.57	0.05	0.96
	6	1,591	2.55	1.20	1,591	2.55	1.20	93.34	5.53	1.13	0.08	0.96
	7	1,591	1.66	1.20	1,591	1.68	1.21	94.22	4.65	1.13	0.07	0.96
	9	1,591	2.32	1.24	1,591	2.32	1.24	96.79	2.89	0.31	0.04	0.99
	10	1,591	2.38	1.25	1,591	2.35	1.25	93.09	5.78	1.13	0.09	0.96
	12	1,591	2.27	1.20	1,591	2.26	1.21	93.40	5.78	0.82	0.08	0.96
Mathematics	1	1,585	1.82	1.27	1,585	1.83	1.27	97.10	2.46	0.44	0.03	0.99
	3	1,585	1.43	0.70	1,585	1.42	0.71	96.34	3.28	0.38	0.04	0.94
	4	1,585	2.41	1.38	1,585	2.41	1.38	96.59	2.59	0.82	0.05	0.98
	6	1,585	2.95	1.27	1,585	2.95	1.27	97.29	2.33	0.38	0.03	0.98
	7	1,585	1.46	0.95	1,585	1.47	0.96	97.10	2.46	0.44	0.04	0.97
	9	1,585	2.89	1.07	1,585	2.88	1.07	96.97	2.59	0.44	0.04	0.98
	10	1,585	2.68	1.05	1,585	2.68	1.05	93.12	6.44	0.44	0.08	0.96
	12	1,585	2.77	1.41	1,585	2.80	1.40	96.53	2.78	0.69	0.05	0.97
Science	1	420	2.50	0.94	420	2.50	0.94	96.43	3.33	0.24	0.04	0.98
	3	420	2.99	1.09	420	2.96	1.10	95.00	4.05	0.95	0.06	0.96
	4	420	2.61	1.01	420	2.61	1.01	95.95	3.57	0.48	0.05	0.97
	6	420	3.10	1.08	420	3.11	1.07	95.48	3.33	1.19	0.07	0.94
	7	420	2.22	1.25	420	2.22	1.25	95.71	4.05	0.24	0.05	0.98
	9	420	3.03	0.85	420	3.03	0.86	96.67	3.10	0.24	0.04	0.97
	10	420	3.09	0.99	420	3.07	1.02	95.48	4.52	0.00	0.05	0.98
	12	420	2.66	1.08	420	2.66	1.09	95.71	3.81	0.48	0.05	0.96

\* 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		
<b>English– Language Arts</b>	1	1,128	2.20	1.22	1,128	2.21	1.22	95.12	3.72	1.15	0.06	0.97
	3	1,128	3.10	0.90	1,128	3.09	0.91	95.74	3.90	0.35	0.05	0.96
	4	1,128	2.91	1.17	1,128	2.90	1.18	91.40	6.74	1.86	0.11	0.93
	6	1,128	2.48	1.10	1,128	2.46	1.10	90.69	7.62	1.68	0.11	0.94
	7	1,128	3.08	1.09	1,128	3.07	1.09	90.69	7.71	1.60	0.12	0.92
	9	1,128	2.52	1.17	1,128	2.51	1.17	90.96	8.16	0.89	0.10	0.96
	10	1,128	2.01	1.20	1,128	2.00	1.21	94.41	4.79	0.80	0.07	0.97
	12	1,128	2.23	1.08	1,128	2.23	1.09	91.05	7.62	1.33	0.11	0.93
<b>Mathematics</b>	1	1,122	2.09	1.25	1,122	2.08	1.25	95.10	4.10	0.80	0.06	0.97
	3	1,122	2.81	1.12	1,122	2.81	1.12	96.26	3.03	0.71	0.05	0.97
	4	1,122	2.59	1.28	1,122	2.59	1.31	93.58	4.37	2.05	0.10	0.94
	6	1,122	2.76	1.39	1,122	2.76	1.39	96.17	2.85	0.98	0.06	0.97
	7	1,122	2.20	1.10	1,122	2.21	1.10	94.74	4.46	0.80	0.07	0.96
	9	1,122	2.16	1.25	1,122	2.16	1.26	95.28	4.01	0.71	0.06	0.97
	10	1,122	2.90	1.30	1,122	2.90	1.30	97.33	2.23	0.45	0.03	0.98
	12	1,122	2.65	1.32	1,122	2.67	1.31	95.54	3.03	1.43	0.08	0.94
<b>Science</b>	1	384	1.86	0.89	384	1.88	0.90	94.01	4.95	1.04	0.07	0.93
	3	384	3.33	0.94	384	3.33	0.94	97.40	2.08	0.52	0.04	0.95
	4	384	2.15	1.15	384	2.18	1.15	92.71	6.77	0.52	0.08	0.97
	6	384	2.05	1.13	384	2.09	1.13	93.49	5.73	0.78	0.08	0.96
	7	384	2.43	1.07	384	2.44	1.06	93.23	5.73	1.04	0.08	0.95
	9	384	2.89	0.97	384	2.88	0.97	95.05	4.17	0.78	0.06	0.95
	10	384	3.34	0.98	384	3.34	0.98	97.40	1.82	0.78	0.04	0.95
	12	384	2.02	1.07	384	2.04	1.07	93.23	5.73	1.04	0.08	0.95

\* Mean absolute difference between first and second ratings

† Pearson correlation between first and second ratings

## Appendix 8.D—IRT Analyses

Table 8.D.1 Score Conversions: Level I, ELA

Raw Score	Freq. Distrib.	Theta	Scale Score	CSEM	Performance Level
40	1,834	N/A	60	0	
39	663	1.1813	54	6	
38	391	0.8180	50	6	
37	309	0.6341	48	5	
36	1,020	0.5104	47	4	
35	497	0.4160	46	3	
34	328	0.3387	45	3	
33	266	0.2725	44	3	
32	749	0.2140	43	3	<b>Advanced</b>
31	400	0.1610	43	3	
30	344	0.1121	42	2	
29	266	0.0663	42	2	
28	662	0.0229	41	2	
27	345	-0.0187	41	2	
26	283	-0.0590	40	2	
25	243	-0.0983	40	2	
24	589	-0.1370	39	2	
23	351	-0.1754	39	2	
22	287	-0.2137	38	2	
21	234	-0.2522	38	2	
20	511	-0.2913	38	2	
19	342	-0.3311	37	2	<b>Proficient</b>
18	249	-0.3722	37	2	
17	245	-0.4148	36	2	
16	408	-0.4594	36	2	
15	274	-0.5068	35	3	
14	233	-0.5576	35	3	
13	195	-0.6130	34	3	
12	271	-0.6744	33	3	
11	231	-0.7442	32	3	<b>Basic</b>
10	162	-0.8259	31	3	
9	150	-0.9251	30	4	
8	294	-1.0518	29	4	
7	156	-1.2232	27	5	
6	158	-1.4656	24	6	<b>Below Basic</b>
5	166	-1.7967	20	7	
4	113	-2.2017	16	4	
3	100	-2.6687	15	1	
2	108	-3.2371	15	0	
1	132	-4.0752	15	0	<b>Far Below Basic</b>
0	363	N/A	15	0	

**Table 8.D.2 Score Conversions: Level II, ELA**

<b>Raw Score</b>	<b>Freq. Distrib.</b>	<b>Theta</b>	<b>Scale Score</b>	<b>CSEM</b>	<b>Performance Level</b>
32	27	N/A	60	0	
31	32	3.4637	56	4	
30	47	2.7047	52	4	
29	79	2.2378	50	3	
28	134	1.8902	48	3	
27	145	1.6071	46	3	
26	189	1.3641	45	2	<b>Advanced</b>
25	228	1.1480	44	2	
24	262	0.9511	43	2	
23	284	0.7683	42	2	
22	334	0.5959	42	2	
21	366	0.4312	41	2	
20	352	0.2715	40	2	
19	385	0.1145	39	2	
18	363	-0.0422	38	2	
17	379	-0.2012	38	2	<b>Proficient</b>
16	328	-0.3652	37	2	
15	392	-0.5375	36	2	
14	312	-0.7213	35	2	
13	289	-0.9196	34	2	
12	218	-1.1348	33	2	
11	175	-1.3679	32	2	<b>Basic</b>
10	117	-1.6187	31	2	
9	78	-1.8875	30	3	
8	73	-2.1758	28	3	
7	58	-2.4864	27	3	
6	50	-2.8223	25	3	<b>Below Basic</b>
5	29	-3.1870	23	3	
4	33	-3.5886	21	3	
3	40	-4.0472	19	4	
2	17	-4.6140	16	2	
1	22	-5.4581	15	1	<b>Far Below Basic</b>
0	35	N/A	15	0	

**Table 8.D.3 Score Conversions: Level III, ELA**

<b>Raw Score</b>	<b>Freq. Distrib.</b>	<b>Theta</b>	<b>Scale Score</b>	<b>CSEM</b>	<b>Performance Level</b>
32	76	N/A	60	0	
31	135	3.3658	52	5	
30	199	2.5945	49	3	
29	238	2.1141	47	3	
28	271	1.7554	46	2	
27	337	1.4649	44	2	
26	353	1.2177	43	2	<b>Advanced</b>
25	388	1.0000	43	2	
24	391	0.8026	42	2	
23	400	0.6193	41	2	
22	369	0.4457	41	2	
21	361	0.2781	40	2	
20	346	0.1138	39	2	
19	344	-0.0493	39	2	
18	310	-0.2132	38	2	
17	285	-0.3793	37	2	
16	292	-0.5490	37	2	<b>Proficient</b>
15	286	-0.7239	36	2	
14	267	-0.9056	35	2	
13	242	-1.0962	35	2	
12	230	-1.2985	34	2	
11	196	-1.5161	33	2	
10	149	-1.7523	32	2	<b>Basic</b>
9	122	-2.0101	31	2	
8	104	-2.2910	30	2	
7	50	-2.5951	29	2	
6	46	-2.9230	28	2	
5	32	-3.2783	26	2	<b>Below Basic</b>
4	36	-3.6706	25	2	
3	29	-4.1220	23	3	
2	29	-4.6850	21	3	
1	19	-5.5292	18	3	<b>Far Below Basic</b>
0	36	N/A	15	0	

**Table 8.D.4 Score Conversions: Level IV, ELA**

<b>Raw Score</b>	<b>Freq. Distrib.</b>	<b>Theta</b>	<b>Scale Score</b>	<b>CSEM</b>	<b>Performance Level</b>
32	113	N/A	60	0	
31	148	2.8727	56	4	
30	237	2.1843	52	5	
29	296	1.7916	50	3	
28	317	1.5157	48	3	
27	348	1.2995	47	3	<b>Advanced</b>
26	415	1.1180	46	2	
25	441	0.9581	45	2	
24	498	0.8125	44	2	
23	424	0.6766	43	2	
22	457	0.5470	42	2	
21	470	0.4216	42	2	
20	454	0.2982	41	2	
19	450	0.1751	40	2	<b>Proficient</b>
18	447	0.0506	40	2	
17	419	-0.0772	39	2	
16	444	-0.2105	38	2	
15	387	-0.3519	37	2	
14	402	-0.5047	36	2	
13	371	-0.6732	35	2	
12	400	-0.8634	34	3	
11	386	-1.0837	33	3	<b>Basic</b>
10	437	-1.3452	31	3	
9	393	-1.6618	30	3	
8	429	-2.0431	27	4	
7	97	-2.4809	25	4	<b>Below Basic</b>
6	85	-2.9479	22	4	
5	81	-3.4236	19	4	
4	65	-3.9118	16	3	<b>Far Below Basic</b>
3	62	-4.4381	15	1	
2	65	-5.0607	15	0	
1	40	-5.9535	15	0	
0	56	N/A	15	0	

Table 8.D.5 Score Conversions: Level V, ELA

Raw Score	Freq. Distrib.	Theta	Scale Score	CSEM	Performance Level
32	213	N/A	60	0	
31	234	3.3078	51	6	
30	353	2.4936	47	3	
29	423	1.9841	46	3	
28	512	1.6084	44	2	<b>Advanced</b>
27	565	1.3092	43	2	
26	584	1.0585	42	2	
25	575	0.8397	41	2	
24	578	0.6423	41	2	
23	553	0.4593	40	2	
22	571	0.2859	39	2	
21	538	0.1184	39	2	
20	470	-0.0461	38	2	<b>Proficient</b>
19	453	-0.2097	37	2	
18	415	-0.3745	37	2	
17	383	-0.5424	36	2	
16	378	-0.7151	35	2	
15	376	-0.8949	35	2	
14	327	-1.0840	34	2	
13	331	-1.2850	33	2	
12	307	-1.5009	32	2	<b>Basic</b>
11	278	-1.7347	32	2	
10	262	-1.9889	31	2	
9	154	-2.2654	30	2	
8	127	-2.5650	28	2	
7	67	-2.8876	27	2	
6	51	-3.2338	26	2	<b>Below Basic</b>
5	43	-3.6065	25	2	
4	50	-4.0150	23	2	
3	46	-4.4805	21	3	
2	48	-5.0553	19	3	<b>Far Below Basic</b>
1	25	-5.9084	16	2	
0	78	N/A	15	0	

**Table 8.D.6 Score Conversions: Level I, Mathematics**

<b>Raw Score</b>	<b>Freq. Distrib.</b>	<b>Theta</b>	<b>Scale Score</b>	<b>CSEM</b>	<b>Performance Level</b>
40	840	N/A	60	0	
39	365	1.5449	50	9	
38	299	1.1864	46	6	
37	264	1.0042	44	4	
36	905	0.8812	43	3	
35	396	0.7871	42	3	<b>Advanced</b>
34	331	0.7100	41	3	
33	242	0.6438	41	2	
32	945	0.5851	40	2	
31	477	0.5318	40	2	
30	334	0.4826	39	2	
29	284	0.4364	39	2	
28	918	0.3925	38	2	
27	416	0.3504	38	2	
26	377	0.3095	37	2	
25	252	0.2694	37	2	
24	739	0.2300	37	2	<b>Proficient</b>
23	375	0.1908	36	2	
22	297	0.1515	36	2	
21	220	0.1119	35	2	
20	637	0.0717	35	2	
19	311	0.0305	35	2	
18	276	-0.0121	34	2	
17	245	-0.0565	34	2	
16	490	-0.1034	33	2	
15	283	-0.1534	33	2	<b>Basic</b>
14	234	-0.2076	32	2	
13	211	-0.2673	32	3	
12	342	-0.3346	31	3	
11	285	-0.4127	30	3	
10	180	-0.5068	29	3	
9	171	-0.6262	28	4	
8	455	-0.7885	27	5	<b>Below Basic</b>
7	184	-1.0257	24	5	
6	148	-1.3703	21	6	
5	156	-1.7961	17	4	
4	133	-2.2535	15	1	
3	133	-2.7457	15	0	
2	143	-3.3278	15	0	<b>Far Below Basic</b>
1	117	-4.1749	15	0	
0	456	N/A	15	0	

**Table 8.D.7 Score Conversions: Level II, Mathematics**

<b>Raw Score</b>	<b>Freq. Distrib.</b>	<b>Theta</b>	<b>Scale Score</b>	<b>CSEM</b>	<b>Performance Level</b>
32	41	N/A	60	0	
31	53	3.1002	60	3	
30	96	2.3022	54	6	
29	194	1.8432	51	5	
28	355	1.4911	48	4	<b>Advanced</b>
27	337	1.1960	46	4	
26	298	0.9485	44	3	
25	298	0.7399	43	3	
24	288	0.5584	41	3	
23	329	0.3942	40	3	
22	336	0.2399	39	3	
21	298	0.0904	38	3	
20	326	-0.0578	37	3	<b>Proficient</b>
19	307	-0.2069	36	3	
18	303	-0.3586	35	3	
17	286	-0.5138	34	3	
16	262	-0.6739	32	3	<b>Basic</b>
15	251	-0.8402	31	3	
14	233	-1.0150	30	3	
13	222	-1.2011	29	3	
12	183	-1.4023	27	3	
11	136	-1.6232	25	4	
10	116	-1.8693	24	4	<b>Below Basic</b>
9	80	-2.1469	22	4	
8	78	-2.4608	19	4	
7	29	-2.8114	17	4	
6	24	-3.1929	15	1	
5	12	-3.5988	15	0	
4	12	-4.0316	15	0	
3	23	-4.5112	15	0	<b>Far Below Basic</b>
2	19	-5.0912	15	0	
1	7	-5.9437	15	0	
0	32	N/A	15	0	

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

<b>Raw Score</b>	<b>Freq. Distrib.</b>	<b>Theta</b>	<b>Scale Score</b>	<b>CSEM</b>	<b>Performance Level</b>
32	46	N/A	60	0	
31	84	2.9374	51	6	
30	165	2.0752	47	4	
29	209	1.5767	44	3	<b>Advanced</b>
28	275	1.2393	43	3	
27	328	0.9869	41	2	
26	374	0.7840	40	2	
25	408	0.6120	40	2	
24	426	0.4601	39	2	
23	357	0.3214	38	2	
22	360	0.1913	38	2	<b>Proficient</b>
21	332	0.0667	37	2	
20	340	-0.0551	36	2	
19	375	-0.1760	36	2	
18	354	-0.2980	35	2	
17	338	-0.4229	34	2	
16	331	-0.5528	34	2	<b>Basic</b>
15	298	-0.6902	33	2	
14	332	-0.8385	32	2	
13	260	-1.0023	32	2	
12	223	-1.1878	31	2	
11	205	-1.4035	30	2	
10	151	-1.6594	28	3	
9	92	-1.9657	27	3	<b>Below Basic</b>
8	95	-2.3261	25	3	
7	44	-2.7302	23	3	
6	28	-3.1564	21	3	
5	28	-3.5901	19	3	
4	19	-4.0359	16	3	
3	17	-4.5188	15	1	<b>Far Below Basic</b>
2	13	-5.0959	15	0	
1	9	-5.9411	15	0	
0	36	N/A	15	0	

**Table 8.D.9 Score Conversions: Level IV, Mathematics**

<b>Raw Score</b>	<b>Freq. Distrib.</b>	<b>Theta</b>	<b>Scale Score</b>	<b>CSEM</b>	<b>Performance Level</b>
32	70	N/A	60	0	
31	83	2.8102	56	5	
30	134	2.2284	52	5	
29	206	1.8654	49	4	
28	224	1.5935	47	3	<b>Advanced</b>
27	348	1.3717	46	3	
26	403	1.1803	44	3	
25	462	1.0084	43	3	
24	500	0.8499	42	3	
23	582	0.7014	41	3	
22	589	0.5611	40	3	
21	602	0.4280	39	2	
20	573	0.3009	38	2	<b>Proficient</b>
19	491	0.1787	37	2	
18	428	0.0597	37	2	
17	408	-0.0582	36	2	
16	398	-0.1773	35	2	
15	371	-0.3005	34	2	
14	389	-0.4314	33	3	<b>Basic</b>
13	466	-0.5747	32	3	
12	461	-0.7374	31	3	
11	533	-0.9296	30	3	
10	410	-1.1670	28	4	
9	310	-1.4705	26	4	<b>Below Basic</b>
8	294	-1.8570	23	4	
7	77	-2.3118	20	5	
6	62	-2.7894	17	4	
5	57	-3.2614	15	1	
4	37	-3.7338	15	0	<b>Far Below Basic</b>
3	31	-4.2356	15	0	
2	27	-4.8269	15	0	
1	17	-5.6834	15	0	
0	60	N/A	15	0	

**Table 8.D.10 Score Conversions: Level V, Mathematics**

<b>Raw Score</b>	<b>Freq. Distrib.</b>	<b>Theta</b>	<b>Scale Score</b>	<b>CSEM</b>	<b>Performance Level</b>
32	350	N/A	60	0	<b>Advanced</b>
31	362	2.1545	50	7	
30	343	1.5924	46	4	
29	411	1.2776	44	3	
28	445	1.0552	43	3	
27	460	0.8796	42	2	
26	516	0.7318	41	2	
25	482	0.6017	40	2	
24	472	0.4836	40	2	
23	524	0.3738	39	2	<b>Proficient</b>
22	491	0.2697	38	2	
21	442	0.1694	38	2	
20	444	0.0712	37	2	
19	432	-0.0265	36	2	
18	384	-0.1251	36	2	
17	419	-0.2265	35	2	
16	394	-0.3327	35	2	
15	391	-0.4463	34	2	
14	362	-0.5709	33	2	
13	312	-0.7120	32	2	
12	349	-0.8782	31	3	
11	324	-1.0834	30	3	
10	323	-1.3515	28	3	<b>Below Basic</b>
9	297	-1.7155	26	4	
8	245	-2.1858	23	4	
7	73	-2.7036	20	4	
6	42	-3.2037	17	4	
5	45	-3.6757	15	2	
4	32	-4.1401	15	0	
3	35	-4.6319	15	0	
2	20	-5.2131	15	0	
1	14	-6.0600	15	0	
0	89	N/A	15	0	

**Table 8.D.11 Score Conversions: Level I, Science**

<b>Raw Score</b>	<b>Freq. Distrib.</b>	<b>Theta</b>	<b>Scale Score</b>	<b>CSEM</b>	<b>Performance Level</b>
40	414	N/A	60	0	
39	117	1.5450	50	9	
38	87	1.1760	46	6	
37	67	0.9891	44	4	
36	264	0.8635	43	3	
35	108	0.7679	42	3	<b>Advanced</b>
34	96	0.6897	41	3	
33	66	0.6229	41	3	
32	186	0.5638	40	2	
31	91	0.5105	39	2	
30	75	0.4612	39	2	
29	53	0.4152	38	2	
28	167	0.3715	38	2	
27	88	0.3297	38	2	
26	78	0.2891	37	2	
25	67	0.2495	37	2	<b>Proficient</b>
24	155	0.2105	36	2	
23	84	0.1717	36	2	
22	75	0.1329	36	2	
21	45	0.0936	35	2	
20	148	0.0537	35	2	
19	66	0.0127	34	2	
18	56	-0.0299	34	2	
17	52	-0.0745	34	2	
16	147	-0.1218	33	2	
15	54	-0.1726	33	2	<b>Basic</b>
14	66	-0.2282	32	2	
13	47	-0.2900	31	3	
12	118	-0.3607	31	3	
11	45	-0.4440	30	3	
10	45	-0.5465	29	3	
9	41	-0.6800	28	4	
8	153	-0.8675	26	5	<b>Below Basic</b>
7	35	-1.1467	23	6	
6	45	-1.5357	19	6	
5	38	-1.9807	15	3	
4	31	-2.4409	15	0	
3	33	-2.9317	15	0	
2	45	-3.5115	15	0	<b>Far Below Basic</b>
1	33	-4.3565	15	0	
0	119	N/A	15	0	

**Table 8.D.12 Score Conversions: Level III, Science**

<b>Raw Score</b>	<b>Freq. Distrib.</b>	<b>Theta</b>	<b>Scale Score</b>	<b>CSEM</b>	<b>Performance Level</b>
32	32	N/A	60	0	<b>Advanced</b>
31	45	2.9363	46	8	
30	74	2.1811	44	3	
29	110	1.7168	42	2	
28	145	1.3718	41	2	
27	167	1.0914	40	2	
26	186	0.8507	39	2	
25	212	0.6363	38	2	
24	249	0.4398	38	2	
23	278	0.2559	37	1	
22	268	0.0809	36	1	
21	241	-0.0881	36	1	
20	266	-0.2532	35	1	
19	226	-0.4162	35	1	
18	190	-0.5787	34	1	<b>Basic</b>
17	167	-0.7422	34	1	
16	131	-0.9083	33	1	
15	114	-1.0785	32	1	
14	95	-1.2546	32	1	
13	81	-1.4385	31	2	
12	61	-1.6328	30	2	
11	42	-1.8400	30	2	
10	40	-2.0632	29	2	<b>Below Basic</b>
9	23	-2.3054	28	2	
8	25	-2.5688	27	2	
7	16	-2.8553	26	2	
6	10	-3.1662	25	2	
5	16	-3.5047	24	2	
4	10	-3.8796	23	2	
3	11	-4.3116	21	2	<b>Far Below Basic</b>
2	4	-4.8524	19	3	
1	3	-5.6714	16	2	
0	13	N/A	15	0	

**Table 8.D.13 Score Conversions: Level IV, Science**

<b>Raw Score</b>	<b>Freq. Distrib.</b>	<b>Theta</b>	<b>Scale Score</b>	<b>CSEM</b>	<b>Performance Level</b>
32	50	N/A	60	0	<b>Advanced</b>
31	75	2.6704	46	8	
30	110	1.9559	43	3	
29	158	1.5289	41	2	
28	190	1.2183	40	2	
27	203	0.9695	39	2	
26	197	0.7580	38	2	<b>Proficient</b>
25	184	0.5704	38	2	
24	181	0.3988	37	2	
23	198	0.2379	36	2	
22	200	0.0841	36	2	
21	210	-0.0653	35	1	
20	191	-0.2127	35	1	<b>Basic</b>
19	158	-0.3601	34	1	
18	163	-0.5094	33	2	
17	131	-0.6625	33	2	
16	110	-0.8216	32	2	
15	107	-0.9889	32	2	
14	85	-1.1666	31	2	<b>Below Basic</b>
13	65	-1.3574	30	2	
12	65	-1.5635	29	2	
11	61	-1.7874	29	2	
10	38	-2.0306	28	2	
9	32	-2.2943	27	2	
8	45	-2.5787	25	2	<b>Far Below Basic</b>
7	7	-2.8840	24	2	
6	9	-3.2116	23	2	
5	8	-3.5661	22	2	
4	16	-3.9584	20	3	
3	4	-4.4120	18	3	
2	9	-4.9815	16	2	<b>Far Below Basic</b>
1	4	-5.8391	15	1	
0	26	N/A	15	0	

**Table 8.D.14 Score Conversions: Level V, Science**

<b>Raw Score</b>	<b>Freq. Distrib.</b>	<b>Theta</b>	<b>Scale Score</b>	<b>CSEM</b>	<b>Performance Level</b>
32	46	N/A	60	0	
31	50	3.3858	45	8	
30	55	2.6934	43	3	
29	78	2.2847	42	2	<b>Advanced</b>
28	110	1.9870	41	2	
27	139	1.7463	40	2	
26	162	1.5379	39	2	
25	178	1.3488	39	1	
24	200	1.1709	38	1	
23	228	0.9988	37	1	<b>Proficient</b>
22	289	0.8289	37	1	
21	227	0.6586	36	1	
20	219	0.4861	36	1	
19	238	0.3108	35	1	
18	213	0.1324	34	1	
17	196	-0.0487	34	1	<b>Basic</b>
16	152	-0.2323	33	1	
15	130	-0.4187	33	1	
14	110	-0.6095	32	1	
13	89	-0.8074	31	2	
12	74	-1.0167	31	2	
11	54	-1.2435	30	2	
10	45	-1.4956	29	2	<b>Below Basic</b>
9	26	-1.7809	28	2	
8	52	-2.1050	27	2	
7	17	-2.4657	26	2	
6	10	-2.8536	24	2	
5	7	-3.2611	23	2	
4	5	-3.6922	21	2	
3	11	-4.1685	20	2	<b>Far Below Basic</b>
2	8	-4.7442	18	3	
1	5	-5.5913	15	1	
0	27	N/A	15	0	

## Appendix 8.E—Disability Distributions

**Table 8.E.1 CAPA Primary Disability Distributions: Level I**

Disability	ELA		Mathematics		Science	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Mental retardation/Intellectual disability	5,572	37.3%	5,561	37.4%	1,475	38.8%
Hard of hearing	83	0.6%	83	0.6%	26	0.7%
Deafness	39	0.3%	39	0.3%	12	0.3%
Speech or language impairment*	112	0.8%	112	0.8%	—	—
Visual impairment	220	1.5%	218	1.5%	86	2.3%
Emotional disturbance	33	0.2%	32	0.2%	11	0.3%
Orthopedic impairment	1,737	11.6%	1,726	11.6%	478	12.6%
Other health impairment	419	2.8%	417	2.8%	86	2.3%
Specific learning disability	86	0.6%	84	0.6%	20	0.5%
Deaf-blindness*	20	0.1%	19	0.1%	—	—
Multiple disabilities	1,807	12.1%	1,796	12.1%	502	13.2%
Autism	4,343	29.1%	4,329	29.1%	1,003	26.4%
Traumatic brain injury	103	0.7%	102	0.7%	30	0.8%
Unknown	348	2.3%	348	2.3%	59	1.6%
<b>TOTAL</b>	<b>14,922</b>	<b>100.0%</b>	<b>14,866</b>	<b>100.0%</b>	<b>3,800</b>	<b>100.0%</b>

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

**Table 8.E.2 CAPA Primary Disability Distributions: Level II**

Disability	ELA		Mathematics	
	Frequency	Percent	Frequency	Percent
Mental retardation/Intellectual disability	1,870	31.8%	1,866	31.8%
Hard of hearing	36	0.6%	35	0.6%
Deafness	29	0.5%	29	0.5%
Speech or language impairment	429	7.3%	428	7.3%
Visual impairment	24	0.4%	23	0.4%
Emotional disturbance	26	0.4%	26	0.4%
Orthopedic impairment	189	3.2%	189	3.2%
Other health impairment	320	5.4%	321	5.5%
Specific learning disability	377	6.4%	377	6.4%
Deaf-blindness*	—	—	—	—
Multiple disabilities	115	2.0%	115	2.0%
Autism	2,244	38.2%	2,243	38.3%
Traumatic brain injury	32	0.5%	32	0.5%
Unknown	180	3.1%	179	3.1%
<b>TOTAL</b>	<b>5,872</b>	<b>100.0%</b>	<b>5,864</b>	<b>100.0%</b>

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

**Table 8.E.3 CAPA Primary Disability Distributions: Level III**

Disability	ELA		Mathematics		Science	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Mental retardation/Intellectual disability	2,430	34.9%	2,427	34.9%	1,281	36.1%
Hard of hearing	51	0.7%	51	0.7%	24	0.7%
Deafness	43	0.6%	43	0.6%	19	0.5%
Speech or language impairment	358	5.1%	358	5.1%	177	5.0%
Visual impairment	32	0.5%	32	0.5%	15	0.4%
Emotional disturbance	46	0.7%	46	0.7%	22	0.6%
Orthopedic impairment	222	3.2%	220	3.2%	115	3.2%
Other health impairment	377	5.4%	375	5.4%	187	5.3%
Specific learning disability	586	8.4%	586	8.4%	309	8.7%
Deaf-blindness*	—	—	—	—	—	—
Multiple disabilities	147	2.1%	147	2.1%	78	2.2%
Autism	2,494	35.8%	2,487	35.8%	1,233	34.7%
Traumatic brain injury	41	0.6%	41	0.6%	15	0.4%
Unknown	132	1.9%	131	1.9%	70	2.0%
<b>TOTAL</b>	<b>6,968</b>	<b>100.0%</b>	<b>6,952</b>	<b>100.0%</b>	<b>3,551</b>	<b>100.0%</b>

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

**Table 8.E.4 CAPA Primary Disability Distributions: Level IV**

Disability	ELA		Mathematics		Science	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Mental retardation/Intellectual disability	4,098	40.4%	4,081	40.4%	1,388	42.2%
Hard of hearing	68	0.7%	68	0.7%	28	0.9%
Deafness	84	0.8%	83	0.8%	27	0.8%
Speech or language impairment	266	2.6%	266	2.6%	67	2.0%
Visual impairment	54	0.5%	54	0.5%	21	0.6%
Emotional disturbance	63	0.6%	62	0.6%	21	0.6%
Orthopedic impairment	444	4.4%	442	4.4%	166	5.0%
Other health impairment	445	4.4%	443	4.4%	140	4.3%
Specific learning disability	768	7.6%	768	7.6%	259	7.9%
Deaf-blindness*	—	—	—	—	—	—
Multiple disabilities	320	3.2%	320	3.2%	118	3.6%
Autism	3,270	32.3%	3,261	32.3%	1,000	30.4%
Traumatic brain injury	67	0.7%	68	0.7%	20	0.6%
Unknown	183	1.8%	183	1.8%	34	1.0%
<b>TOTAL</b>	<b>10,134</b>	<b>100.0%</b>	<b>10,103</b>	<b>100.0%</b>	<b>3,290</b>	<b>100.0%</b>

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

**Table 8.E.5 CAPA Primary Disability Distributions: Level V**

Disability	ELA		Mathematics		Science	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Mental retardation/Intellectual disability	4,468	43.1%	4,455	43.2%	1,491	43.2%
Hard of hearing	65	0.6%	65	0.6%	19	0.6%
Deafness	96	0.9%	95	0.9%	33	1.0%
Speech or language impairment	177	1.7%	176	1.7%	68	2.0%
Visual impairment	57	0.5%	57	0.6%	25	0.7%
Emotional disturbance	145	1.4%	142	1.4%	45	1.3%
Orthopedic impairment	458	4.4%	456	4.4%	141	4.1%
Other health impairment	461	4.4%	465	4.5%	159	4.6%
Specific learning disability	839	8.1%	836	8.1%	293	8.5%
Deaf-blindness*	–	–	–	–	–	–
Multiple disabilities	446	4.3%	443	4.3%	153	4.4%
Autism	2,793	26.9%	2,778	26.9%	915	26.5%
Traumatic brain injury	66	0.6%	66	0.6%	26	0.8%
Unknown	293	2.8%	286	2.8%	82	2.4%
<b>TOTAL</b>	<b>10,368</b>	<b>100.0%</b>	<b>10,324</b>	<b>100.0%</b>	<b>3,450</b>	<b>100.0%</b>

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

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

### Quality Control of Task Development

The task development process for the CAPA prior to the 2014 administration is described in detail in Chapter 3, starting on page 17. The next sections highlight elements of the process devoted specifically to the quality control of the tasks that were previously developed and reused during the 2014 CAPA administration.

#### Task Specifications

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

#### Task Writers

The tasks for the CAPA were written by task writers with a thorough understanding of the California content standards. The task writers were carefully screened and selected by senior ETS content staff and approved by the CDE. Only those with strong content and teaching backgrounds, experienced with students who have severe cognitive disabilities, were invited to participate in an extensive training program for task writers.

#### Internal Contractor Reviews

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

#### Content Review

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

The artwork and graphics for the tasks were created during the internal content review period so assessment specialists could evaluate the correctness and appropriateness of the art early in the task development process. ETS selected visuals that were relevant to the

task content and that were easily understood so students would not struggle to determine the purpose or meaning of the questions.

### **Editorial Review**

Another step in the ETS internal review process involved a team of specially trained editors who checked tasks 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 included rounds of copyediting and proofreading. ETS 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, conducted this bias and sensitivity review. These staff members had been trained to identify and eliminate tasks that contained content that could be construed as offensive to, or biased against, members of specific ethnic, racial, or gender groups.

### **Assessment Director Review**

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

### **Assessment Review Panel Review**

The ARPs were panels that advised the CDE and ETS on areas related to task development for the CAPA. The ARPs were responsible for reviewing all newly developed tasks for alignment to the California content standards. The ARPs also reviewed 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 was responsible for reviewing and approving the achievement tests that were used statewide for the testing of students in California public schools in grades two through eleven. The SPAR panel representatives ensured that the CAPA tasks conformed 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-tested newly developed tasks to obtain statistical information about task performance. This information was used to evaluate tasks that were candidates for use in operational test forms. These tasks that were flagged after field-test and operational use were examined carefully at data review meetings, where content experts discussed tasks that had poor statistics and did not meet the psychometric criteria for task quality. The CDE defined the criteria for acceptable or unacceptable task statistics. These criteria ensured that the task (1) had an appropriate level of difficulty for the target population; (2) discriminated well between examinees who differ in ability; and (3) conformed well to the statistical model underlying the measurement of the intended constructs. The results of analyses for differential item functioning (DIF) were used to make judgments about the appropriateness of items for various subgroups when the items were first used.

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

The field-test items that appeared in the CAPA administered in 2014 were statistically reviewed in data review meetings in 2013, the year they were originally administered. There was no data review of field-test items in 2014.

## Quality Control of the Item Bank

After the data review, tasks were placed in the item bank along with their statistics and reviewers' evaluations of their quality. ETS then delivered 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 are very important. The ETS internal item bank database resides on a server within the ETS firewall; access to the SQL Server database is strictly controlled by means of system administration. The electronic item banking application includes a login/password system to authorize access to the database or designated portions of the database. In addition, only users authorized to access the specific database are able to use the item bank. Users are authorized by a designated administrator at the CDE and at ETS.

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

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

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

## Quality Control of Test Form Development

The ETS Assessment Development group is committed to providing the highest quality product to the students of California and has in place a number of quality control (QC) checks to ensure that outcome. During the task development process, there were multiple senior reviews of tasks, including one by the assessment director. Test forms certification was a formal quality control process established as a final checkpoint prior to printing. In it, content, editorial, and senior development staff review test forms for accuracy and clueing issues.

ETS also included quality checks throughout preparation of the form planners. A form planner specifications document was developed by the test development team lead with input from ETS's item bank and statistics groups; this document was then reviewed by all team members who built forms at a training session specific to form planners before the

form-building process started. After trained content team members signed off on a form planner, a representative from the internal QC group reviewed each file for accuracy against the specifications document. Assessment directors reviewed and signed off on form planners prior to processing.

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

## Quality Control of Test Materials

### Collecting Test Materials

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

All scorable materials are delivered to the Pearson scanning and scoring facilities in Iowa City, Iowa. The nonscorable materials, including *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 California Technical Assistance Center (CalTAC) at ETS monitors returns and notifies LEAs that do not return their materials in a timely manner. CalTAC contacts the LEA CAASPP coordinators and works with them to facilitate the return of the test materials.

### Processing Test Materials

Upon receipt of the 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 its facilities. The materials are removed carefully from the shipping cartons and examined for a number of conditions, including physical damage, shipping errors, and omissions. A visual inspection to compare the number of students recorded on the School and Grade Identification (SGID) 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. LEAs are contacted by phone if there are any missing shipments or the quantity of materials returned appears to be less than expected.

## Quality Control of Scanning

Before any CAASPP documents are scanned, Pearson conducts a complete check of the scanning system. ETS and Pearson create test decks for every test and form. Each test deck consists of approximately 25 answer documents marked to cover response ranges, demographic data, blanks, double marks, and other responses. Fictitious students are created to verify that each marking possibility is processed correctly by the scanning program. The output file generated as a result of this activity is thoroughly checked against each answer document after each stage to verify that the scanner is capturing marks correctly. When the program output is confirmed to match the expected results, a scan program release form is signed and the scan program is placed in the production environment under configuration management.

The intensity levels of each scanner are constantly monitored for quality control purposes. Intensity diagnostics sheets are run before and during each batch to verify that the scanner is working properly. In the event that a scanner fails to properly pick up 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.

## **Quality Control of Image Editing**

Prior to submitting any CAASPP operational documents through the image editing process, Pearson creates a mock set of documents to test all of the errors listed in the edit specifications. The set of test documents is used to verify that each image of the document is saved so that an editor will be able to review the documents through an interactive interface. The edits are confirmed to show the appropriate error, the correct image to edit the 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 against expected results to ensure the appropriate corrections are made. The post file will have all keyed corrections and any defaults from the edit specifications.

## **Quality Control of Answer Document Processing and Scoring**

### **Accountability of Answer Documents**

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

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

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

### **Processing of Answer Documents**

Prior to processing operational answer documents and executing subsequent data processing programs, ETS conducts an end-to-end test. As part of this test, ETS prepares approximately 700 test cases covering all tests and many scenarios designed to exercise particular business rule logic. ETS marks answer documents for those 700 test cases. They are then scanned, scored, and aggregated. The results at various inspection points are checked by psychometricians and Data Quality Services staff. Additionally, a post-scan test file of approximately 50,000 records across the CAASPP System is scored and aggregated

to test a broader range of scoring and aggregation scenarios. These procedures assure that students and LEAs receive the correct scores when the actual scoring process is carried out. In 2014, end-to-end testing also included the inspection of results in electronic reporting.

### **Scoring and Reporting Specifications**

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

- General Reporting Specifications
- Form Planner Specifications
- Aggregation Rules
- “What If” List
- Edit Specifications (which include matching information from observer documents to examiner documents for 10 percent of the CAPA that is administered)

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.

### **Storing Answer Documents**

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

## **Quality Control of Psychometric Processes**

### **Quality Control of Task (Item) Analyses and the Scoring Process**

When the forms were first administered in 2013, psychometric analyses conducted at ETS underwent comprehensive quality checks by a team of psychometricians and data analysts. Detailed checklists were consulted by members of the team for each of the statistical procedures performed on each CAPA following its original administration. Quality assurance checks also included a comparison of the current year’s statistics to statistics from previous years. The results of preliminary classical task analyses that provided a check on scoring reasonableness and the application of scoring rubrics were also reviewed by a senior psychometrician. The tasks that were flagged for questionable statistical attributes were sent to test development staff for their review; their comments were reviewed by the psychometricians before tasks were approved to be included in the equating process.

The results of the equating process were reviewed by a psychometric manager in addition to the aforementioned team of psychometricians and data analysts. If the senior psychometrician and the manager reached a consensus that an equating result did not conform to the norm, special binders were prepared for review by senior psychometric advisors at ETS, along with several pieces of informative analyses to facilitate the process.

When the forms were equated following their original administration, a few additional checks are performed for each process as described below.

## Calibrations

During the calibration that was conducted for the original administration of each form and that is described in more detail in Chapter 2 starting on page 13, checks were made to ascertain that the correct options for the analyses were selected. Checks were 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 performed detailed reviews of plots and statistics to investigate if the model fit the data.

## Scaling

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

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

## Scoring Tables

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

- All raw scores were included in the tables;
- Scale scores increased as raw scores increased;
- The minimum reported scale score was 15 and the maximum reported scale score was 60; and
- The cut points for the performance levels were correctly identified.

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

## Score Verification Process

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

## Year-to-Year Comparison Analyses

Year-to-year comparison analyses are conducted each year for quality control of the scoring procedure in general and as reasonableness checks for the 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 LEAs.

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

### **Offloads to Test Development**

During the original administration of the CAPA forms that were reused in 2014, the statistics based on classical task analyses were obtained to ensure the stability of the statistics. The resulting classical statistics for all items were provided to test development staff in specially designed Excel spreadsheets called “statistical offloads.” The offloads were thoroughly checked by the psychometric staff before their release for test development review.

## **Quality Control of Reporting**

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

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

All reports are required to include a single, accurate CDS code, a charter school number (if applicable), an LEA name, and a school name. All elements conform to the CDE’s official CDS code and naming records. From the start of processing through scoring and reporting, the CDS Master File is used to verify and confirm accurate codes and names. The CDS Master File is provided by the CDE to ETS throughout the year as updates are available.

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

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

### **Electronic Reporting**

Because no equating was conducted during the 2014 administration, students’ scale scores and performance levels for the CAPA were made available to LEAs prior to the printing of paper reports. The Quick-turnaround Reporting module of the Test Management System made it possible for LEAs to securely download an electronic reporting file containing these results.

Before an LEA could download a student data file, ETS statisticians approved a QC file of test results data and ETS IT successfully processed it. Once the data were deemed reliable and Pearson processed a scorable answer document for every student who took the CAPA

in that test administration for the LEA, the LEA was notified that these results were available.

### **Excluding Student Scores from Summary Reports**

ETS provides specifications to the CDE that document when to exclude student scores from summary reports. These specifications include the logic for handling answer documents that, for example, indicate the student 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 three most recent years of administration—2012, 2013, and 2014—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 the 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 were first administered in 2003. Subsequently, the CAPA were revised to better link them 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 then 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 on page 156 contains the number of examinees assessed and the means and standard deviations of examinees' scale scores in the base year (2009) and in 2012, 2013, and 2014 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 156; please note that this information may differ slightly from information found on the CDE's CAASPP reporting Web page at <http://caaspp.cde.ca.gov> due to differing dates on which data were accessed. The goal is for all students to achieve at or above the proficient level by 2014.

Table 10.A.3 through Table 10.A.5 show for each CAPA the distribution of scale scores observed in the base year, in 2012, 2013, and 2014. 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 tests, 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 (task) and test analysis results of the CAPA over the comparison years indicate that the CAPA meet the technical criteria established in professional standards for high-stakes tests. In addition, every year, efforts were made to improve the technical quality of each CAPA.

Table 10.B.1 and Table 10.B.2 in Appendix 10.B, which start on page 160, present, respectively, the average task scores and the equated IRT *b*-value means for the tasks in each CAPA. The average task scores were affected by both the difficulty of the items and the abilities of the students administered the tasks.

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, Examinee Performance

**Table 10.A.1 Number of Examinees Tested, Scale Score Means, and Standard Deviations of CAPA Across Base Year (2009), 2012, 2013, and 2014**

Content Area	CAPA	Number of Examinees (valid scores)				Scale Score Mean and Standard Deviation							
		Base	2012	2013	2014	Base Mean S.D.	2012 Mean S.D.	2013 Mean S.D.	2014 Mean S.D.				
English– Language Arts	I	12,531	14,098	14,707	14,922	40.84	12.02	40.76	11.04	41.76	10.60	41.45	10.85
	II	6,587	6,668	6,383	5,872	39.24	7.46	38.82	6.91	38.56	6.04	38.29	6.24
	III	6,614	7,105	7,160	6,968	39.12	5.94	39.56	6.46	39.51	5.82	39.44	5.98
	IV	9,853	10,091	10,261	10,134	39.19	7.75	39.02	8.45	39.16	8.16	38.95	8.18
	V	10,517	10,424	10,678	10,368	38.54	6.21	38.72	6.04	38.87	6.35	38.66	6.56
Mathematics	I	12,484	14,065	14,673	14,866	35.11	9.74	36.15	9.00	36.57	9.22	36.61	9.52
	II	6,569	6,650	6,381	5,864	37.60	9.56	37.28	8.50	37.46	8.55	37.25	8.66
	III	6,602	7,094	7,142	6,952	36.58	6.64	36.34	5.54	36.44	5.72	36.26	5.86
	IV	9,831	10,068	10,241	10,103	36.41	8.80	37.14	7.50	36.79	7.55	36.55	7.67
	V	10,485	10,392	10,644	10,324	37.51	8.85	37.49	8.08	37.41	7.91	37.27	7.95
Science	I	3,296	3,564	3,724	3,800	35.59	11.25	36.25	10.25	37.35	10.29	37.61	11.14
	III	3,267	3,556	3,446	3,551	36.24	5.45	36.33	4.65	36.10	4.63	36.09	4.65
	IV	3,190	3,299	3,275	3,290	35.56	5.53	36.02	4.98	35.91	5.37	35.73	5.69
	V	3,396	3,424	3,435	3,450	35.35	5.34	36.22	5.21	35.84	4.98	35.86	5.12

**Table 10.A.2 Percentage of Proficient and Above and Percentage of Advanced Across Base Year (2009), 2012, 2013, and 2014**

Content Area	CAPA	% Proficient and Above				% Advanced			
		Base	2012	2013	2014	Base	2012	2013	2014
English–Language Arts	I	75%	81%	83%	83%	51%	59%	58%	58%
	II	78%	80%	80%	79%	41%	43%	44%	42%
	III	83%	81%	86%	84%	42%	54%	52%	50%
	IV	77%	72%	75%	74%	37%	40%	42%	41%
	V	80%	80%	81%	79%	42%	44%	47%	44%
Mathematics	I	61%	67%	69%	69%	29%	34%	38%	38%
	II	62%	65%	67%	65%	33%	35%	35%	33%
	III	65%	71%	65%	64%	31%	20%	28%	27%
	IV	60%	66%	66%	65%	31%	27%	30%	30%
	V	67%	69%	72%	71%	34%	33%	39%	37%
Science	I	59%	64%	68%	66%	33%	34%	39%	41%
	III	69%	71%	71%	70%	19%	18%	17%	16%
	IV	58%	66%	66%	66%	15%	14%	17%	18%
	V	61%	70%	66%	65%	17%	23%	24%	24%

**Table 10.A.3 Observed Score Distributions of CAPA Across Base Year (2009), 2012, 2013, and 2014 for ELA**

Observed Score Distributions	Level I				Level II				Level III				Level IV				Level V			
	Base	2012	2013	2014	Base	2012	2013	2014	Base	2012	2013	2014	Base	2012	2013	2014	Base	2012	2013	2014
60	2,230	1,554	1,883	1,834	405	53	33	27	199	71	72	76	219	131	113	113	274	173	189	213
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	N/A	N/A	N/A	N/A	N/A
54–56	N/A	473	632	663	N/A	83	41	32	N/A	N/A	N/A	N/A	239	208	179	148	N/A	N/A	N/A	N/A
51–53	624	352	N/A	N/A	N/A	142	41	47	N/A	149	98	135	N/A	235	257	237	N/A	N/A	253	234
48–50	388	276	727	700	375	192	218	213	304	178	186	199	653	692	632	613	400	247	N/A	N/A
45–47	299	1,551	1,752	1,845	375	771	395	334	426	973	550	509	967	1,334	1,243	1,204	517	777	836	776
42–44	1,708	1,876	1,987	2,025	795	960	1,265	1,108	934	1,305	1,599	1,469	1,534	1,415	1,883	1,849	1,277	1,556	1,754	1,661
39–41	1,784	2,843	2,502	2,473	1,090	1,294	1,226	1,103	1,341	1,788	1,947	1,820	1,911	1,669	1,798	1,770	3,097	2,902	3,101	2,815
36–38	1,567	1,940	2,223	2,276	1,776	1,483	1,547	1,462	2,044	1,068	1,168	1,173	1,669	1,220	1,192	1,233	2,179	2,241	1,762	1,721
33–35	1,559	975	1,018	973	1,081	926	897	819	891	874	910	935	1,008	1,178	1,251	1,157	1,698	1,364	1,339	1,412
30–32	694	570	597	543	362	292	386	370	258	255	353	375	822	887	832	830	572	672	940	1,001
27–29	545	405	450	450	154	182	135	131	111	212	86	96	398	358	339	429	211	197	190	194
24–26	140	154	117	158	89	99	39	50	45	84	82	68	83	310	111	97	113	130	89	94
21–23	128	126	N/A	N/A	28	81	66	62	34	38	50	58	70	81	51	85	59	43	90	96
18–20	128	N/A	144	166	12	31	32	40	5	22	20	19	125	137	86	81	33	28	37	48
15–17	737	1,003	675	816	45	79	62	74	22	88	39	36	155	236	294	288	87	94	98	103

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

**Table 10.A.4 Observed Score Distributions of CAPA Across Base Year (2009), 2012, 2013, and 2014 for Mathematics**

Observed Score Distributions	Level I				Level II				Level III				Level IV				Level V			
	Base	2012	2013	2014	Base	2012	2013	2014	Base	2012	2013	2014	Base	2012	2013	2014	Base	2012	2013	2014
60	603	641	739	840	417	71	112	94	134	68	37	46	269	93	91	70	767	529	362	350
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	N/A	N/A	N/A	N/A	N/A
54–56	N/A	N/A	N/A	N/A	N/A	107	116	96	N/A	N/A	N/A	N/A	N/A	130	84	83	N/A	N/A	N/A	N/A
51–53	N/A	N/A	N/A	N/A	386	147	184	194	N/A	N/A	103	84	391	268	125	134	N/A	N/A	N/A	N/A
48–50	237	282	345	365	N/A	321	385	355	230	94	N/A	N/A	295	361	194	206	N/A	356	351	362
45–47	382	228	303	299	338	631	357	337	N/A	98	178	165	687	380	587	572	499	404	401	343
42–44	934	1,263	1,570	1,565	682	749	710	596	762	533	499	484	689	925	1,399	1,365	1,104	1,063	1,453	1,316
39–41	1,465	2,302	2,645	2,613	886	909	1,057	953	1,274	1,429	1,586	1,536	1,436	2,429	1,851	1,773	1,804	2,134	2,090	1,994
36–38	2,775	3,619	3,368	3,374	1,049	1,131	1,007	931	1,579	2,043	1,930	1,764	1,687	1,664	2,058	1,900	2,475	2,296	2,232	2,193
33–35	2,628	2,611	2,443	2,462	1,053	830	653	589	1,105	1,405	1,310	1,321	1,229	1,486	1,197	1,158	1,524	1,425	1,502	1,566
30–32	1,053	1,061	1,068	1,072	658	564	802	746	837	856	1,026	1,020	1,319	1,163	1,349	1,460	918	1,060	1,029	985
27–29	407	673	802	806	547	495	411	405	320	296	229	243	888	549	389	410	473	276	308	323
24–26	492	195	171	184	137	354	269	252	200	101	68	95	286	193	301	310	278	554	321	297
21–23	174	146	169	148	209	103	95	80	39	56	57	72	257	157	262	294	321	N/A	256	245
18–20	177	N/A	N/A	N/A	34	53	74	78	33	18	19	28	75	54	91	77	61	59	64	73
15–17	1,157	1,044	1,050	1,138	173	185	149	158	89	97	100	94	323	216	263	291	261	236	275	277

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

**Table 10.A.5 Observed Score Distributions of CAPA Across Base Year (2009), 2012, 2013, and 2014 for Science**

Observed Score Distributions	Level I				Level III				Level IV				Level V			
	Base	2012	2013	2014	Base	2012	2013	2014	Base	2012	2013	2014	Base	2012	2013	2014
60	280	272	322	414	69	28	19	32	46	48	50	50	33	58	38	46
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	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	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	N/A	N/A	N/A	N/A
48–50	81	65	123	117	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
45–47	69	79	80	87	105	64	55	45	44	65	79	75	46	74	50	50
42–44	267	339	403	439	122	221	188	184	157	98	107	110	129	104	137	133
39–41	394	452	518	514	493	535	521	498	393	445	496	551	373	547	588	589
36–38	588	828	846	767	934	1,280	1,224	1,248	1,010	1,113	1,003	960	1,288	1,186	1,217	1,163
33–35	611	656	609	568	1,093	1,011	885	980	864	1,073	927	853	874	878	852	929
30–32	271	262	272	276	268	265	363	393	420	292	376	367	332	369	335	327
27–29	108	83	92	86	104	68	105	88	155	83	129	196	196	139	135	123
24–26	207	193	131	153	29	37	37	42	36	39	56	52	36	19	19	27
21–23	N/A	N/A	48	35	20	17	22	21	10	10	17	17	25	19	20	12
18–20	49	41	32	45	10	3	1	4	19	11	13	20	14	8	13	19
15–17	371	294	248	299	20	27	26	16	36	22	22	39	50	23	31	32

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

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

**Table 10.B.1 Average Item Score of CAPA Operational Test Tasks Across Base Year (2009), 2012, 2013, and 2014**

Content Area	Level	Average Item Score			
		Base	2012	2013	2014
English–Language Arts	I	3.37	3.12	3.25	3.24
	II	2.91	2.38	2.30	2.27
	III	2.91	2.52	2.51	2.49
	IV	2.51	2.33	2.30	2.28
	V	2.73	2.57	2.61	2.57
Mathematics	I	2.70	2.86	2.96	2.98
	II	2.70	2.45	2.52	2.51
	III	2.70	2.39	2.51	2.48
	IV	2.37	2.49	2.31	2.30
	V	2.76	2.65	2.55	2.53
Science	I	2.75	2.91	3.04	3.07
	III	2.71	2.60	2.63	2.63
	IV	2.47	2.69	2.69	2.68
	V	2.47	2.74	2.53	2.53

**Table 10.B.2 Mean IRT *b*-values for Operational Test Tasks Across Base Year (2009), 2012, 2013, and 2014**

Content Area	Level	Mean IRT <i>b</i> -value			
		Base	2012	2013	2014
English–Language Arts	I	–0.74	–0.60	–0.56	–0.56
	II	–1.54	–0.76	–0.65	–0.65
	III	–1.52	–0.82	–0.78	–0.78
	IV	–0.93	–0.87	–0.75	–0.75
	V	–1.19	–0.91	–0.99	–0.99
Mathematics	I	–0.29	–0.24	–0.27	–0.27
	II	–1.18	–0.96	–1.00	–1.00
	III	–1.29	–0.93	–1.00	–1.00
	IV	–0.85	–0.81	–0.66	–0.66
	V	–1.21	–1.09	–0.99	–0.99
Science	I	–0.23	–0.31	–0.32	–0.32
	III	–1.29	–1.05	–1.10	–1.10
	IV	–0.95	–1.11	–1.14	–1.14
	V	–0.54	–0.65	–0.57	–0.57

**Table 10.B.3 Mean Polyserial Correlation of CAPA Operational Test Tasks Across Base Year (2009), 2012, 2013, and 2014**

Content Area	Level	Mean Polyserial Correlation			
		Base	2012	2013	2014
English–Language Arts	I	0.81	0.80	0.78	0.79
	II	0.75	0.78	0.74	0.74
	III	0.75	0.80	0.78	0.79
	IV	0.78	0.80	0.79	0.79
	V	0.79	0.79	0.80	0.81
Mathematics	I	0.79	0.76	0.77	0.77
	II	0.78	0.77	0.75	0.76
	III	0.76	0.71	0.73	0.73
	IV	0.79	0.73	0.74	0.75
	V	0.78	0.77	0.77	0.78
Science	I	0.82	0.79	0.79	0.81
	III	0.75	0.72	0.74	0.73
	IV	0.75	0.70	0.74	0.75
	V	0.78	0.76	0.75	0.76

**Table 10.B.4 Score Reliabilities and SEM of CAPA Across Base Year (2009), 2012, 2013, and 2014**

Content Area	Level	Reliability				SEM			
		Base	2012	2013	2014	Base	2012	2013	2014
English–Language Arts	I	0.91	0.89	0.88	0.88	3.67	3.89	3.92	3.92
	II	0.84	0.87	0.84	0.84	2.49	2.32	2.38	2.39
	III	0.86	0.90	0.88	0.89	2.26	2.17	2.28	2.27
	IV	0.88	0.90	0.89	0.89	2.50	2.33	2.43	2.43
	V	0.89	0.89	0.90	0.90	2.35	2.27	2.20	2.22
Mathematics	I	0.87	0.85	0.86	0.86	4.00	4.19	4.11	4.11
	II	0.88	0.86	0.85	0.85	2.58	2.60	2.45	2.45
	III	0.87	0.81	0.83	0.84	2.54	2.67	2.59	2.57
	IV	0.88	0.83	0.83	0.84	2.62	2.66	2.65	2.63
	V	0.87	0.86	0.87	0.87	2.70	2.76	2.67	2.64
Science	I	0.91	0.88	0.88	0.90	3.76	4.03	3.97	3.83
	III	0.85	0.84	0.85	0.84	2.43	2.45	2.27	2.27
	IV	0.85	0.81	0.85	0.86	2.46	2.47	2.35	2.35
	V	0.87	0.85	0.85	0.85	2.30	2.27	2.32	2.33