

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



# California Assessment of Student Performance and Progress California Science Test 2023–24 Technical Report

**Final Submitted February 25, 2025**

**By ETS**



**Contract #CN220002**

## Table of Contents

[Chapter 1: Introduction 1](#_Toc190780445)

[1.1. Background 1](#_Toc190780446)

[1.2. Purpose of the Assessment 2](#_Toc190780447)

[1.3. Assessment Design 3](#_Toc190780448)

[1.4. Intended Population 3](#_Toc190780449)

[1.5. Intended Use and Purpose of Test Scores 3](#_Toc190780450)

[1.6. Testing Window 4](#_Toc190780451)

[1.7. Significant CAST Developments in 2023–24 4](#_Toc190780452)

[1.7.1. Updated Statewide Testing Window 4](#_Toc190780453)

[1.7.2. Student Score Reports Redesign 5](#_Toc190780454)

[1.7.3. Test Delivery 5](#_Toc190780455)

[1.7.4. Accessibility Resources 5](#_Toc190780456)

[1.8. Groups and Organizations Involved with the CAASPP System 5](#_Toc190780457)

[1.8.1. California State Board of Education 5](#_Toc190780458)

[1.8.2. California Department of Education 6](#_Toc190780459)

[1.8.3. California Educators 6](#_Toc190780460)

[1.8.4. National Science Experts 6](#_Toc190780461)

[1.8.5. Contractors 6](#_Toc190780462)

[1.9. Systems Overview and Functionality 7](#_Toc190780463)

[1.9.1. Test Operations Management System 7](#_Toc190780464)

[1.9.2. Test Delivery System 8](#_Toc190780465)

[1.9.3. Practice and Training Tests 8](#_Toc190780466)

[1.9.4. California Educator Reporting System 9](#_Toc190780467)

[1.9.5. Test Results for California’s Assessments Website 9](#_Toc190780468)

[1.9.6. Constructed-Response Scoring Systems for ETS 9](#_Toc190780469)

[1.10. Overview of the Technical Report 11](#_Toc190780470)

[References 12](#_Toc190780471)

[Chapter 2: Overview of CAST Processes 13](#_Toc190780472)

[2.1. Item Development 13](#_Toc190780473)

[2.1.1. Design Guidelines 13](#_Toc190780474)

[2.1.2. Content Guidelines 13](#_Toc190780475)

[2.1.3. Item Types Guidelines 13](#_Toc190780476)

[2.2. Test Assembly 14](#_Toc190780477)

[2.2.1. Test Blueprint 14](#_Toc190780478)

[2.3. Test Administration 14](#_Toc190780479)

[2.3.1. Test Security and Confidentiality 15](#_Toc190780480)

[2.3.2. Procedures to Maintain Standardization 15](#_Toc190780481)

[2.4. Fairness and Accessibility 16](#_Toc190780482)

[2.4.1. Overview 16](#_Toc190780483)

[2.4.2. Student Accessibility Resources 16](#_Toc190780484)

[2.4.3. Description of Differential Item Functioning 17](#_Toc190780485)

[2.5. Scores 17](#_Toc190780486)

[2.5.1. Score Reporting 17](#_Toc190780487)

[2.5.2. Aggregation Procedures 17](#_Toc190780488)

[2.6. Psychometric Analyses 18](#_Toc190780489)

[2.6.1. Description of Classical Item Analyses 18](#_Toc190780490)

[2.6.2. Description of Item Response Theory Analyses 18](#_Toc190780491)

[References 19](#_Toc190780492)

[Chapter 3: Item Development and Review 20](#_Toc190780493)

[3.1. Use of Evidence-Centered Design 20](#_Toc190780494)

[3.1.1. Principles 20](#_Toc190780495)

[3.1.2. Theory of Action Model 20](#_Toc190780496)

[3.1.3. Incorporation into Item Development Processes 22](#_Toc190780497)

[3.1.4. Item Types and Features 23](#_Toc190780498)

[3.2. Guidelines 24](#_Toc190780499)

[3.2.1. Plan 25](#_Toc190780500)

[3.2.2. Process 25](#_Toc190780501)

[3.2.3. Item Specifications 26](#_Toc190780502)

[3.2.4. Recruitment and Selection of Item Writers 26](#_Toc190780503)

[3.2.5. Item Writer Training 26](#_Toc190780504)

[3.3. ETS Item Review Process 27](#_Toc190780505)

[3.3.1. Overview 27](#_Toc190780506)

[3.3.2. ETS Content Review 28](#_Toc190780507)

[3.3.3. ETS Accessibility Review 28](#_Toc190780508)

[3.3.4. ETS Editorial Review 28](#_Toc190780509)

[3.3.5. ETS Sensitivity and Fairness Review 28](#_Toc190780510)

[3.4. California Department of Education Review 29](#_Toc190780511)

[3.5. California Educator Review 29](#_Toc190780512)

[3.5.1. California Educators as Content Experts 29](#_Toc190780513)

[3.5.2. Composition of Item Review Panels 29](#_Toc190780514)

[3.5.3. Meetings for Review of CAST Field Test Items 30](#_Toc190780515)

[3.5.4. Data Review Meeting 31](#_Toc190780516)

[References 33](#_Toc190780517)

[Chapter 4: Test Assembly 34](#_Toc190780518)

[4.1. Overview 34](#_Toc190780519)

[4.1.1. Design 34](#_Toc190780520)

[4.2. Automated Test Assembly 35](#_Toc190780521)

[4.3. Test Blueprint and Other Content Specifications 35](#_Toc190780522)

[4.3.1. Test Blueprint 35](#_Toc190780523)

[4.4. Test Production Process 36](#_Toc190780524)

[4.4.1. Content Criteria 36](#_Toc190780525)

[4.4.2. Test Forms 37](#_Toc190780526)

[4.4.3. Psychometric Review 37](#_Toc190780527)

[4.4.4. Content Review of Forms 40](#_Toc190780528)

[4.4.5. California Department of Education Forms Review 40](#_Toc190780529)

[4.4.6. Configuration of the Test Delivery System 41](#_Toc190780530)

[4.4.7. Test Form Delivery 41](#_Toc190780531)

[4.5. Performance Expectation Coverage 42](#_Toc190780532)

[4.6. Special Version Forms 42](#_Toc190780533)

[4.6.1. Braille Form 42](#_Toc190780534)

[4.6.2. Paper–Pencil Form 43](#_Toc190780535)

[4.6.3. Forms with Accessibility Features Other Than Braille 43](#_Toc190780536)

[References 44](#_Toc190780537)

[Appendix 4.A: Test Blueprint 45](#_Toc190780538)

[Appendix 4.B: Performance Expectation Distribution for Segment A 48](#_Toc190780539)

[Blueprint 48](#_Toc190780540)

[Accessibility Information 52](#_Toc190780541)

[Chapter 5: Test Administration 65](#_Toc190780542)

[5.1. Overview 65](#_Toc190780543)

[5.1.1. In-Person and Remote Testing 65](#_Toc190780544)

[5.1.2. Student Test-Taking Requirements 65](#_Toc190780545)

[5.2. User Roles and Standardization 66](#_Toc190780546)

[5.2.1. Local Educational Agency CAASPP Coordinator 66](#_Toc190780547)

[5.2.2. Site CAASPP Coordinator 67](#_Toc190780548)

[5.2.3. Test Administrator 67](#_Toc190780549)

[5.2.4. Instructions for Test Administration 68](#_Toc190780550)

[5.3. Local Educational Agency Training 69](#_Toc190780551)

[5.3.1. Synchronous and Asynchronous Training 70](#_Toc190780552)

[5.3.2. Videos and Guides 70](#_Toc190780553)

[5.3.3. Training for Proper Identification and Assignment of Designated Supports and Accommodations 71](#_Toc190780554)

[5.3.4. Feedback for Continuous Improvement Survey 72](#_Toc190780555)

[5.4. Accessibility Resources 72](#_Toc190780556)

[5.4.1. Accessibility Resource Categories 73](#_Toc190780557)

[5.4.2. Identification and Selection 77](#_Toc190780558)

[5.4.3. Assignment 78](#_Toc190780559)

[5.4.4. Delivery of Embedded and Non-Embedded Resources to Students 79](#_Toc190780560)

[5.4.5. Usage of Designated Supports and Accommodations 79](#_Toc190780561)

[5.5. Practice and Training Tests 83](#_Toc190780562)

[5.6. Test Security and Confidentiality 83](#_Toc190780563)

[5.6.1. The ETS Office of Testing Integrity 84](#_Toc190780564)

[5.6.2. Procedures to Maintain Standardization of Test Security 84](#_Toc190780565)

[5.6.3. Test Security Monitoring 85](#_Toc190780566)

[5.6.4. Security of Electronic Files Using a Firewall 86](#_Toc190780567)

[5.6.5. Transfer of Scores via Secure Data Exchange 86](#_Toc190780568)

[5.6.6. Data Management in the Secure Database 86](#_Toc190780569)

[5.6.7. Statistical Analysis on Secure Servers 87](#_Toc190780570)

[5.6.8. Student Confidentiality 87](#_Toc190780571)

[5.6.9. Student Test Results 87](#_Toc190780572)

[5.6.10. Security and Test Administration Incident Reporting System Process 88](#_Toc190780573)

[5.6.11. Appeals 90](#_Toc190780574)

[References 93](#_Toc190780575)

[Appendix 5.A: Accessibility Resource Assignment 95](#_Toc190780576)

[Chapter 6: Standard Setting 99](#_Toc190780577)

[Reference 100](#_Toc190780578)

[Chapter 7: Scoring and Reporting 101](#_Toc190780579)

[7.1. CAST Scoring Process 101](#_Toc190780580)

[7.1.1. Scoring Constructed-Response Items 101](#_Toc190780581)

[7.1.2. Scoring for Selected-Response Items 111](#_Toc190780582)

[7.2. Student Test Scores 111](#_Toc190780583)

[7.2.1. Theta Scores 112](#_Toc190780584)

[7.2.2. Scale Scores for the Total Assessment 112](#_Toc190780585)

[7.3. Summary Statistics 113](#_Toc190780586)

[7.3.1. Total-Test Achievement Levels 114](#_Toc190780587)

[7.3.2. Domain Performance Levels 116](#_Toc190780588)

[7.3.3. Demographic Student Group Summaries 120](#_Toc190780589)

[7.4. Reports Produced and Scores for Each Report 121](#_Toc190780590)

[7.4.1. Online Reporting 121](#_Toc190780591)

[7.4.2. Special Cases 122](#_Toc190780592)

[7.4.3. Types of Score Reports 122](#_Toc190780593)

[7.4.4. Score Report Applications 124](#_Toc190780594)

[7.4.5. Criteria for Interpreting Test Scores 125](#_Toc190780595)

[7.4.6. Criteria for Interpreting Score Reports 125](#_Toc190780596)

[References 126](#_Toc190780597)

[Accessibility Information 127](#_Toc190780598)

[Alternative Text for Equation 7.1 127](#_Toc190780599)

[Alternative Text for Equation 7.2 127](#_Toc190780600)

[Alternative Text for Equation 7.3 127](#_Toc190780601)

[Alternative Text for Equation 7.4 127](#_Toc190780602)

[Appendix 7.A: Overall Theta Score Distribution 128](#_Toc190780603)

[Appendix 7.B: Scale Scores of Assessments 134](#_Toc190780604)

[Appendix 7.C: Demographic Student Group Summaries of Participation 152](#_Toc190780605)

[Appendix 7.D: Demographic Student Group Summaries of Overall Achievement Levels 165](#_Toc190780606)

[Appendix 7.E: Demographic Student Group Summaries of Domain Performance   
Levels 189](#_Toc190780607)

[Chapter 8: Psychometric Analyses 225](#_Toc190780608)

[8.1. Overview 225](#_Toc190780609)

[8.1.1. Summary of the Analyses 225](#_Toc190780610)

[8.1.2. Samples Used for Analyses 225](#_Toc190780611)

[8.1.3. Test-Taking Rates 226](#_Toc190780612)

[8.2. Classical Item Analyses 227](#_Toc190780613)

[8.2.1. Classical Item Difficulty Indices (*p*-value) 227](#_Toc190780614)

[8.2.2. Item-Total Correlation 228](#_Toc190780615)

[8.2.3. Distribution of Item Scores 229](#_Toc190780616)

[8.2.4. Omit Rates 229](#_Toc190780617)

[8.2.5. Completion Rates 230](#_Toc190780618)

[8.2.6. Distractor Analyses 230](#_Toc190780619)

[8.2.7. Summary of Classical Item Analyses Flagging Criteria 231](#_Toc190780620)

[8.2.8. Classical Item Analyses Results 232](#_Toc190780621)

[8.3. Differential Item Functioning Analyses 232](#_Toc190780622)

[8.3.1. Differential Item Functioning Procedure for Dichotomous Items 233](#_Toc190780623)

[8.3.2. Differential Item Functioning Procedure for Polytomous Items 234](#_Toc190780624)

[8.3.3. Classification 235](#_Toc190780625)

[8.3.4. Differential Item Functioning Analysis Results 236](#_Toc190780626)

[8.4. Item Response Theory Analyses 237](#_Toc190780627)

[8.4.1. Item Response Theory Model 238](#_Toc190780628)

[8.4.2. Data Preparation 238](#_Toc190780629)

[8.4.3. Equating 239](#_Toc190780630)

[8.4.4. Calibration and Linking for the Field Test Items 239](#_Toc190780631)

[8.4.5. Parameter Estimates 240](#_Toc190780632)

[8.4.6. Scaling the Scores 242](#_Toc190780633)

[8.5. Response Time Analyses 244](#_Toc190780634)

[8.6. Reliability Analyses 248](#_Toc190780635)

[8.6.1. Sample for Reliability Analyses 248](#_Toc190780636)

[8.6.2. Reliability Measures 249](#_Toc190780637)

[8.6.3. Standard Error of Measurement 249](#_Toc190780638)

[8.6.4. Student Group Reliabilities and Standard Errors of Measurement 250](#_Toc190780639)

[8.6.5. Conditional Standard Errors of Measurement 251](#_Toc190780640)

[8.6.6. Decision Classification Analyses 253](#_Toc190780641)

[8.6.7. Interrater Agreement 255](#_Toc190780642)

[8.7. Validity Evidence 259](#_Toc190780643)

[8.7.1. Design of CAST 260](#_Toc190780644)

[8.7.2. Content 261](#_Toc190780645)

[8.7.3. Response Processes 262](#_Toc190780646)

[8.7.4. Internal Structure 263](#_Toc190780647)

[8.7.5. Relations to Other Variables 263](#_Toc190780648)

[8.7.6. Consequences of Testing 265](#_Toc190780649)

[8.8. Special Considerations for the Spanish-Translation Version 265](#_Toc190780650)

[References 267](#_Toc190780651)

[Accessibility Information 270](#_Toc190780652)

[Alternative Text for Equation 8.1 270](#_Toc190780653)

[Alternative Text for Equation 8.2 270](#_Toc190780654)

[Alternative Text for Equation 8.3 270](#_Toc190780655)

[Alternative Text for Equation 8.4 270](#_Toc190780656)

[Alternative Text for Equation 8.5 270](#_Toc190780657)

[Alternative Text for Equation 8.6 270](#_Toc190780658)

[Alternative Text for Equation 8.7 270](#_Toc190780659)

[Alternative Text for Equation 8.8 271](#_Toc190780660)

[Alternative Text for Equation 8.9 271](#_Toc190780661)

[Alternative Text for Equation 8.10 271](#_Toc190780662)

[Alternative Text for Equation 8.11 271](#_Toc190780663)

[Alternative Text for Equation 8.12 271](#_Toc190780664)

[Alternative Text for Equation 8.13 271](#_Toc190780665)

[Alternative Text for Equation 8.14 271](#_Toc190780666)

[Alternative Text for Equation 8.15 271](#_Toc190780667)

[Alternative Text for Equation 8.16 271](#_Toc190780668)

[Alternative Text for Equation 8.17 271](#_Toc190780669)

[Alternative Text for Equation 8.18 272](#_Toc190780670)

[Alternative Text for Equation 8.19 272](#_Toc190780671)

[Alternative Text for Equation 8.20 272](#_Toc190780672)

[Alternative Text for Equation 8.21 272](#_Toc190780673)

[Alternative Text for Equation 8.22 272](#_Toc190780674)

[Alternative Text for Equation 8.23 272](#_Toc190780675)

[Alternative Text for Equation 8.24 272](#_Toc190780676)

[Alternative Text for Equation 8.25 272](#_Toc190780677)

[Alternative Text for Equation 8.26 273](#_Toc190780678)

[Alternative Text for Equation 8.27 273](#_Toc190780679)

[Alternative Text for Equation 8.28 273](#_Toc190780680)

[Alternative Text for Equation 8.29 273](#_Toc190780681)

[Appendix 8.A: Test-Taking Rates 274](#_Toc190780682)

[Appendix 8.B: Item Difficulty Distribution 282](#_Toc190780683)

[Appendix 8.C: Item-Total Correlation Distribution 284](#_Toc190780684)

[Appendix 8.D: Item Discrimination Parameter Distribution 286](#_Toc190780685)

[Appendix 8.E: Item Difficulty Parameter Distribution 292](#_Toc190780686)

[Appendix 8.F: Response Time Analyses 298](#_Toc190780687)

[Appendix 8.G: Reliability Analyses 302](#_Toc190780688)

[Appendix 8.H: Conditional Standard Errors of Measurement 316](#_Toc190780689)

[8.8.1. CSEMs for Grade Five 316](#_Toc190780690)

[8.8.2. CSEMs for Grade Eight 320](#_Toc190780691)

[8.8.3. CSEMs for Grade Ten 324](#_Toc190780692)

[8.8.4. CSEMs for Grade Eleven 328](#_Toc190780693)

[8.8.5. CSEMs for Grade Twelve 332](#_Toc190780694)

[8.8.6. CSEMs for the High School Grade Band 336](#_Toc190780695)

[Appendix 8.I: Analyses of Classification 340](#_Toc190780696)

[Appendix 8.J: Interrater Reliability 344](#_Toc190780697)

[Appendix 8.K: Correlations to Smarter Balanced Test Scores 347](#_Toc190780698)

[Appendix 8.L: Correlations Between CAST Domain Scores and Total Test Scores and Smarter Balanced Composite Claim Scores and Total Test Scores 359](#_Toc190780699)

[Chapter 9: Quality-Control Procedures 362](#_Toc190780700)

[9.1. Quality Control of Item Development 362](#_Toc190780701)

[9.2. Quality Control of Test Assembly and Delivery 362](#_Toc190780702)

[9.3. Quality Control of Test Materials 363](#_Toc190780703)

[9.3.1. Developing Assessments 363](#_Toc190780704)

[9.3.2. Test Administration Manuals 364](#_Toc190780705)

[9.3.3. Collecting Test Materials 364](#_Toc190780706)

[9.3.4. Processing Test Materials 364](#_Toc190780707)

[9.4. Quality Control of Test Administration 365](#_Toc190780708)

[9.5. Quality Control of Scoring 365](#_Toc190780709)

[9.5.1. Machine-Scoring Procedures 365](#_Toc190780710)

[9.5.2. Human Scoring 366](#_Toc190780711)

[9.5.3. Artificial Intelligence Scoring Verification 366](#_Toc190780712)

[9.5.4. Development of Scoring Specifications 366](#_Toc190780713)

[9.5.5. Paper–Pencil Scoring 366](#_Toc190780714)

[9.6. Quality Control of Psychometric Processes 367](#_Toc190780715)

[9.6.1. Scoring Verification 367](#_Toc190780716)

[9.6.2. Psychometric Analyses 367](#_Toc190780717)

[9.7. Quality Control of Reporting 368](#_Toc190780718)

[9.7.1. Exclusion of Student Scores from Summary Reports 368](#_Toc190780719)

[9.8. Quality Control of End-to-End Testing 369](#_Toc190780720)

[9.8.1. Computer-based Assessments 369](#_Toc190780721)

[9.8.2. Paper–Pencil Tests 369](#_Toc190780722)

[References 370](#_Toc190780723)

[Chapter 10: Student Survey 371](#_Toc190780724)

[10.1. Student Survey Questions 371](#_Toc190780725)

[10.2. Student Survey Results 372](#_Toc190780726)

[Appendix 10.A: Results of the Student Survey 373](#_Toc190780727)

[Chapter 11: Continuous and Systematic Improvement 381](#_Toc190780728)

[11.1. 2023–24 Feedback for Continuous Improvement Survey 381](#_Toc190780729)

[11.2. Communications 381](#_Toc190780730)

[11.2.1. Updated Program Website 381](#_Toc190780731)

[11.3. Reporting 381](#_Toc190780732)

[11.4. Test Design 381](#_Toc190780733)

[11.5. Item Development Improvements 382](#_Toc190780734)

[11.5.1. Item Content Specifications 382](#_Toc190780735)

[11.5.2. Processes 382](#_Toc190780736)

[11.5.3. Item Bank 382](#_Toc190780737)

[11.6. Psychometric Analyses 383](#_Toc190780738)

[11.6.1. Updated Artificial Intelligence Scoring Criteria 383](#_Toc190780739)

[11.6.2. Combined Item Analyses 383](#_Toc190780740)

[11.7. Accessibility Resources 384](#_Toc190780741)

[References 385](#_Toc190780742)

### List of Tables

[Acronyms and Initialisms Used in the *California Science Test Technical Report* 15](#_Toc190951747)

[Table 3.1 Item Types Used in CAST 23](#_Toc190951748)

[Table 3.2 Total Number of Items Developed per Grade Level and the Grade Band 25](#_Toc190951749)

[Table 3.3 Item Data Review Results 32](#_Toc190951750)

[Table 4.1 Number of Unique Items Assessed on the CAST 35](#_Toc190951751)

[Table 4.2 The Number of Blocks for Each Segment for 2023–24 CAST General Computer-‍based Forms 37](#_Toc190951752)

[Table 4.3 Desired Item Difficulty Distribution in Segment A 39](#_Toc190951753)

[Table 4.4 Theta Cut Scores Based on Standard Setting Results 39](#_Toc190951754)

[Table 4.5 Number of Blocks Reviewed Across All Forms 40](#_Toc190951755)

[Table 4.6 PEs Assessed on the 2023–24 CAST 42](#_Toc190951756)

[Table 4.7 Cumulative Coverage of PEs Over Three Years 42](#_Toc190951757)

[Table 4.A.1 CAST Blueprint for Segments Contributing to Individual Scores from the Earth and Space Sciences Domain 45](#_Toc190951758)

[Table 4.A.2 CAST Blueprint for Segments Contributing to Individual Scores from the Life Sciences Domain 46](#_Toc190951759)

[Table 4.A.3 CAST Blueprint for Segments Contributing to Individual Scores from the Physical Sciences Domain 46](#_Toc190951760)

[Table 4.A.4 CAST Blueprint for Segments Contributing to Individual Scores for Grade   
Five 47](#_Toc190951761)

[Table 4.A.5 CAST Blueprint for Segments Contributing to Individual Scores for Grade   
Eight 47](#_Toc190951762)

[Table 4.A.6 CAST Blueprint for Segments Contributing to Individual Scores for High   
School 47](#_Toc190951763)

[Table 5.1 Number of Students by Assessment Location 65](#_Toc190951764)

[Table 5.2 Summary of Accommodations and Designated Supports Used by Students 81](#_Toc190951765)

[Table 5.3 Types of Appeals 89](#_Toc190951766)

[Table 5.4 Number and Types of Incidents Submitted in STAIRS 91](#_Toc190951767)

[Table 5.5 Number of Appeals Requested in STAIRS—‍All Grade Levels 92](#_Toc190951768)

[Table 5.A.1 Accessibility Resource Assignment, Grades Five and Eight—All Tested 95](#_Toc190951769)

[Table 5.A.2 Accessibility Resource Assignment, High School Grades—All Tested 97](#_Toc190951770)

[Table 7.1 CAST Sample Selection for Human-Scoring Procedures 103](#_Toc190951771)

[Table 7.2 Summary of Characteristics of Human Raters Scoring CAST 104](#_Toc190951772)

[Table 7.3 Number of Operational CR Items Flagged 108](#_Toc190951773)

[Table 7.4 Performance Criteria for Acceptance for AI Scoring 111](#_Toc190951774)

[Table 7.5 Scaling Constants 112](#_Toc190951775)

[Table 7.6 Required Number of Items for Reporting Domain Scores 113](#_Toc190951776)

[Table 7.7 Mean and SD of Theta Scores and Scale Scores 114](#_Toc190951777)

[Table 7.8 Scale Score Ranges for Achievement Levels 114](#_Toc190951778)

[Table 7.9 Percent of Students in Each Achievement Level for Total Scores 115](#_Toc190951779)

[Table 7.10 Description of Science Domain Performance Levels 116](#_Toc190951780)

[Table 7.11 Scale Score Ranges in Each Performance Level for the Earth and Space Sciences Domain 118](#_Toc190951781)

[Table 7.12 Scale Score Ranges in Each Performance Level for the Life Sciences   
Domain 118](#_Toc190951782)

[Table 7.13 Scale Score Ranges in Each Performance Level for the Physical Sciences Domain 118](#_Toc190951783)

[Table 7.14 Percentage of Students in Each Performance Level for the Earth and Space Sciences Domain 118](#_Toc190951784)

[Table 7.15 Percentage of Students in Each Performance Level for the Life Sciences   
Domain 119](#_Toc190951785)

[Table 7.16 Percentage of Students in Each Performance Level for the Physical Sciences Domain 119](#_Toc190951786)

[Table 7.17 Demographic Student Groups to Be Reported 120](#_Toc190951787)

[Table 7.A.1 Overall Theta Score Distribution for Grade Five 128](#_Toc190951788)

[Table 7.A.2 Overall Theta Score Distribution for Grade Eight 129](#_Toc190951789)

[Table 7.A.3 Overall Theta Score Distribution for Grade Ten 130](#_Toc190951790)

[Table 7.A.4 Overall Theta Score Distribution for Grade Eleven 131](#_Toc190951791)

[Table 7.A.5 Overall Theta Score Distribution for Grade Twelve 132](#_Toc190951792)

[Table 7.A.6 Overall Theta Score Distribution for High School 133](#_Toc190951793)

[Table 7.B.1 CAST Grade Five Scale Scores 134](#_Toc190951794)

[Table 7.B.2 CAST Grade Eight Scale Scores 137](#_Toc190951795)

[Table 7.B.3 CAST Grade Ten Scale Scores 140](#_Toc190951796)

[Table 7.B.4 CAST Grade Eleven Scale Scores 143](#_Toc190951797)

[Table 7.B.5 CAST Grade Twelve Scale Scores 146](#_Toc190951798)

[Table 7.B.6 CAST High School Scale Scores 149](#_Toc190951799)

[Table 7.C.1 Demographic Student Group Summaries for Grade Five 152](#_Toc190951800)

[Table 7.C.2 Demographic Student Group Summaries for Grade Eight 155](#_Toc190951801)

[Table 7.C.3 Demographic Student Group Summaries for Grade Ten 157](#_Toc190951802)

[Table 7.C.4 Demographic Student Group Summaries for Grade Eleven 159](#_Toc190951803)

[Table 7.C.5 Demographic Student Group Summaries for Grade Twelve 161](#_Toc190951804)

[Table 7.C.6 Demographic Student Group Summaries for High School (All Grades   
Tested) 163](#_Toc190951805)

[Table 7.D.1 Percent of Students in Each Achievement Level for Total Scores by Demographic Student Group for Grade Five 165](#_Toc190951806)

[Table 7.D.2 Percent of Students in Each Achievement Level for Total Scores by Demographic Student Group for Grade Eight 169](#_Toc190951807)

[Table 7.D.3 Percent of Students in Each Achievement Level for Total Scores by Demographic Student Group for Grade Ten 173](#_Toc190951808)

[Table 7.D.4 Percent of Students in Each Achievement Level for Total Scores by Demographic Student Group for Grade Eleven 177](#_Toc190951809)

[Table 7.D.5 Percent of Students in Each Achievement Level for Total Scores by Demographic Student Group for Grade Twelve 181](#_Toc190951810)

[Table 7.D.6 Percent of Students in Each Achievement Level for Total Scores by Demographic Student Group for High School 185](#_Toc190951811)

[Table 7.E.1 Percent of Students in Each Performance Level by Demographic Student Group—Earth and Space Sciences Domain, Grade Five 189](#_Toc190951812)

[Table 7.E.2 Percent of Students in Each Performance Level by Demographic Student Group—Earth and Space Sciences Domain, Grade Eight 191](#_Toc190951813)

[Table 7.E.3 Percent of Students in Each Performance Level by Demographic Student Group—Earth and Space Sciences Domain, Grade Ten 193](#_Toc190951814)

[Table 7.E.4 Percent of Students in Each Performance Level by Demographic Student Group—Earth and Space Sciences Domain, Grade Eleven 195](#_Toc190951815)

[Table 7.E.5 Percent of Students in Each Performance Level by Demographic Student Group—Earth and Space Sciences Domain, Grade Twelve 197](#_Toc190951816)

[Table 7.E.6 Percent of Students in Each Performance Level by Demographic Student Group—Earth and Space Sciences Domain, High School 199](#_Toc190951817)

[Table 7.E.7 Percent of Students in Each Performance Level by Demographic Student Group—Life Sciences Domain, Grade Five 201](#_Toc190951818)

[Table 7.E.8 Percent of Students in Each Performance Level by Demographic Student Group—Life Sciences Domain, Grade Eight 203](#_Toc190951819)

[Table 7.E.9 Percent of Students in Each Performance Level by Demographic Student Group—Life Sciences Domain, Grade Ten 205](#_Toc190951820)

[Table 7.E.10 Percent of Students in Each Performance Level by Demographic Student Group—Life Sciences Domain, Grade Eleven 207](#_Toc190951821)

[Table 7.E.11 Percent of Students in Each Performance Level by Demographic Student Group—Life Sciences Domain, Grade Twelve 209](#_Toc190951822)

[Table 7.E.12 Percent of Students in Each Performance Level by Demographic Student Group—Life Sciences Domain, High School 211](#_Toc190951823)

[Table 7.E.13 Percent of Students in Each Performance Level by Demographic Student Group—Physical Sciences Domain, Grade Five 213](#_Toc190951824)

[Table 7.E.14 Percent of Students in Each Performance Level by Demographic Student Group—Physical Sciences Domain, Grade Eight 215](#_Toc190951825)

[Table 7.E.15 Percent of Students in Each Performance Level by Demographic Student Group—Physical Sciences Domain, Grade Ten 217](#_Toc190951826)

[Table 7.E.16 Percent of Students in Each Performance Level by Demographic Student Group—Physical Sciences Domain, Grade Eleven 219](#_Toc190951827)

[Table 7.E.17 Percent of Students in Each Performance Level by Demographic Student Group—Physical Sciences Domain, Grade Twelve 221](#_Toc190951828)

[Table 7.E.18 Percent of Students in Each Performance Level by Demographic Student Group—Physical Sciences Domain, High School 223](#_Toc190951829)

[Table 8.1 CAST Test-Taking Rates of the Full Population 226](#_Toc190951830)

[Table 8.2 Item Difficulty Distributions 232](#_Toc190951831)

[Table 8.3 Item-Total Correlation Distributions 232](#_Toc190951832)

[Table 8.4 Student Groups for DIF Comparison 233](#_Toc190951833)

[Table 8.5 DIF Categories for Dichotomous Items 235](#_Toc190951834)

[Table 8.6 DIF Categories for Polytomous Items 235](#_Toc190951835)

[Table 8.7 Number of Items by DIF Category for Grade Five 236](#_Toc190951836)

[Table 8.8 Number of Items by DIF Category for Grade Eight 236](#_Toc190951837)

[Table 8.9 Number of Items by DIF Category for High School 237](#_Toc190951838)

[Table 8.10 Item Discrimination Parameter Distribution by Grade Level or Grade Band 241](#_Toc190951839)

[Table 8.11 Item Difficulty Parameter Distribution by Grade Level or Grade Band 241](#_Toc190951840)

[Table 8.12 Testing Time (in Minutes) for the Total Assessment 246](#_Toc190951841)

[Table 8.13 Summary Statistics for Scale Scores and Theta Scores, Reliability, and   
SEMs 250](#_Toc190951842)

[Table 8.14 Average CSEMs at Thresholds for Achievement Levels 253](#_Toc190951843)

[Table 8.15 Decision Accuracy for Reaching an Achievement Level 254](#_Toc190951844)

[Table 8.16 Decision Consistency for Reaching an Achievement Level 255](#_Toc190951845)

[Table 8.17 Frequencies of Ratings 256](#_Toc190951846)

[Table 8.18 Correlations Between CAST and ELPAC Scores 264](#_Toc190951847)

[Table 8.19 Spanish Item Analyses Domain Distribution 266](#_Toc190951848)

[Table 8.20 Spanish Item Analyses Item Type Distribution 266](#_Toc190951849)

[Table 8.A.1 CAST Test-Taking Rates for Grade Five by Student Group 274](#_Toc190951850)

[Table 8.A.2 CAST Test-Taking Rates for Grade Eight by Student Group 276](#_Toc190951851)

[Table 8.A.3 CAST Test-Taking Rates for Grade Ten by Student Group 277](#_Toc190951852)

[Table 8.A.4 CAST Test-Taking Rates for Grade Eleven by Student Group 278](#_Toc190951853)

[Table 8.A.5 CAST Test-Taking Rates for Grade Twelve by Student Group 279](#_Toc190951854)

[Table 8.A.6 CAST Test-Taking Rates for High School (All Grades Tested) by Student   
Group 280](#_Toc190951855)

[Table 8.A.7 CAST Test-Taking Rates for the Grade Twelve Cohort by Student Group 281](#_Toc190951856)

[Table 8.B.1 Item Difficulty Distributions by Item Type 282](#_Toc190951857)

[Table 8.B.2 Item Difficulty Distributions by Content Domain 283](#_Toc190951858)

[Table 8.C.1 Item-Total Correlation Distributions by Item Type 284](#_Toc190951859)

[Table 8.C.2 Item-Total Correlation Distributions by Content Domain 285](#_Toc190951860)

[Table 8.D.1 Item Discrimination Parameter Distribution by Item Type for Grade Five 286](#_Toc190951861)

[Table 8.D.2 Item Discrimination Parameter Distribution by Item Type for Grade Eight 287](#_Toc190951862)

[Table 8.D.3 Item Discrimination Parameter Distribution by Item Type for High School 288](#_Toc190951863)

[Table 8.D.4 Item Discrimination Parameter Distribution by Content Domain for Grade   
Five 289](#_Toc190951864)

[Table 8.D.5 Item Discrimination Parameter Distribution by Content Domain for Grade   
Eight 290](#_Toc190951865)

[Table 8.D.6 Item Discrimination Parameter Distribution by Content Domain for High   
School 291](#_Toc190951866)

[Table 8.E.1 Item Difficulty Parameter Distribution by Item Type for Grade Five 292](#_Toc190951867)

[Table 8.E.2 Item Difficulty Parameter Distribution by Item Type for Grade Eight 293](#_Toc190951868)

[Table 8.E.3 Item Difficulty Parameter Distribution by Item Type for High School 294](#_Toc190951869)

[Table 8.E.4 Item Difficulty Parameter Distribution by Content Domain for Grade Five 295](#_Toc190951870)

[Table 8.E.5 Item Difficulty Parameter Distribution by Content Domain for Grade Eight 296](#_Toc190951871)

[Table 8.E.6 Item Difficulty Parameter Distribution by Content Domain for High School 297](#_Toc190951872)

[Table 8.F.1 Testing Time (in Minutes) by Segment 299](#_Toc190951873)

[Table 8.F.2 Testing Time (in Minutes) per Item by Item Type 300](#_Toc190951874)

[Table 8.F.3 Testing Time (in Minutes) for Operational Items by Content Domain 301](#_Toc190951875)

[Table 8.G.1 Reliabilities and SEMs by Demographic Student Groups for Grade Five 302](#_Toc190951876)

[Table 8.G.2 Reliabilities and SEMs by Demographic Student Groups for Grade Eight 304](#_Toc190951877)

[Table 8.G.3 Reliabilities and SEMs by Demographic Student Groups for Grade Ten 306](#_Toc190951878)

[Table 8.G.4 Reliabilities and SEMs by Demographic Student Groups for Grade Eleven 308](#_Toc190951879)

[Table 8.G.5 Reliabilities and SEMs by Demographic Student Groups for Grade Twelve 310](#_Toc190951880)

[Table 8.G.6 Reliabilities and SEMs by Demographic Student Groups for High School 312](#_Toc190951881)

[Table 8.G.7 Reliabilities and SEMs by ELPAC Performance Levels 314](#_Toc190951882)

[Table 8.H.1 Average CSEM by Theta Data, Grade Five 317](#_Toc190951883)

[Table 8.H.2 Average CSEM by Theta Data, Grade Eight 321](#_Toc190951884)

[Table 8.H.3 Average CSEM by Theta Data, Grade Ten 325](#_Toc190951885)

[Table 8.H.4 Average CSEM by Theta Data, Grade Eleven 329](#_Toc190951886)

[Table 8.H.5 Average CSEM by Theta Data, Grade Twelve 333](#_Toc190951887)

[Table 8.H.6 Average CSEM by Theta Data, High School 337](#_Toc190951888)

[Table 8.I.1 Grade Five: Decision Accuracy 340](#_Toc190951889)

[Table 8.I.2 Grade Five: Decision Consistency 341](#_Toc190951890)

[Table 8.I.3 Grade Eight: Decision Accuracy 341](#_Toc190951891)

[Table 8.I.4 Grade Eight: Decision Consistency 341](#_Toc190951892)

[Table 8.I.5 High School—Grade Ten: Decision Accuracy 341](#_Toc190951893)

[Table 8.I.6 High School—Grade Ten: Decision Consistency 342](#_Toc190951894)

[Table 8.I.7 High School—Grade Eleven: Decision Accuracy 342](#_Toc190951895)

[Table 8.I.8 High School—Grade Eleven: Decision Consistency 342](#_Toc190951896)

[Table 8.I.9 High School—Grade Twelve: Decision Accuracy 342](#_Toc190951897)

[Table 8.I.10 High School—Grade Twelve: Decision Consistency 343](#_Toc190951898)

[Table 8.I.11 High School: Decision Accuracy 343](#_Toc190951899)

[Table 8.I.12 High School: Decision Consistency 343](#_Toc190951900)

[Table 8.J.1 Interrater Reliability and Descriptive Statistics for the Ratings by AI and   
Human Raters in AI Scoring of Operational CR Items for Grade Five 344](#_Toc190951901)

[Table 8.J.2 Interrater Reliability and Descriptive Statistics for the Ratings by AI and   
Human Raters in AI Scoring of Operational CR Items for Grade Eight 345](#_Toc190951902)

[Table 8.J.3 Interrater Reliability and Descriptive Statistics for the Ratings by AI and   
Human Raters in AI Scoring of Operational CR Items for High School 346](#_Toc190951903)

[Table 8.K.1 Correlations with Smarter Balanced ELA Test Scores for Grade Five 347](#_Toc190951904)

[Table 8.K.2 Correlations with Smarter Balanced Mathematics Test Scores for Grade   
Five 349](#_Toc190951905)

[Table 8.K.3 Correlations with Smarter Balanced ELA Test Scores for Grade Eight 351](#_Toc190951906)

[Table 8.K.4 Correlations with Smarter Balanced Mathematics Test Scores for Grade   
Eight 353](#_Toc190951907)

[Table 8.K.5 Correlations with Smarter Balanced ELA Test Scores for Grade Eleven 355](#_Toc190951908)

[Table 8.K.6 Correlations with Smarter Balanced Mathematics Test Scores for Grade   
Eleven 357](#_Toc190951909)

[Table 8.L.1 Correlations Between CAST Domain and Total Test Scores and Smarter Balanced ELA Composite Claim and Total Test Scores, Grade Five 359](#_Toc190951910)

[Table 8.L.2 Correlations Between CAST Domain and Total Test Scores and Smarter Balanced Mathematics Composite Claim and Total Test Scores, Grade Five 359](#_Toc190951911)

[Table 8.L.3 Correlations Between CAST Domain and Total Test Scores and Smarter Balanced ELA Composite Claim and Total Test Scores, Grade Eight 360](#_Toc190951912)

[Table 8.L.4 Correlations Between CAST Domain and Total Test Scores and Smarter Balanced Mathematics Composite Claim and Total Test Scores, Grade Eight 360](#_Toc190951913)

[Table 8.L.5 Correlations Between CAST Domain and Total Test Scores and Smarter Balanced ELA Composite Claim and Total Test Scores, Grade Eleven 361](#_Toc190951914)

[Table 8.L.6 Correlations Between CAST Domain and Total Test Scores and Smarter Balanced Mathematics Composite Claim and Total Test Scores, Grade Eleven 361](#_Toc190951915)

[Table 10.A.1 Scale Score Distribution by Responses to Question 1 for Grade Five 373](#_Toc190951916)

[Table 10.A.2 Scale Score Distribution by Responses to Question 2 for Grade Five 373](#_Toc190951917)

[Table 10.A.3 Scale Score Distribution by Responses to Question 3 for Grade Five 373](#_Toc190951918)

[Table 10.A.4 Scale Score Distribution by Responses to Question 1 for Grade Eight 374](#_Toc190951919)

[Table 10.A.5 Scale Score Distribution by Responses to Question 2 for Grade Eight 374](#_Toc190951920)

[Table 10.A.6 Scale Score Distribution by Responses to Question 3 for Grade Eight 374](#_Toc190951921)

[Table 10.A.7 Scale Score Distribution by Responses to Question 1 for Grade Ten 375](#_Toc190951922)

[Table 10.A.8 Scale Score Distribution by Responses to Question 2 for Grade Ten 375](#_Toc190951923)

[Table 10.A.9 Scale Score Distribution by Responses to Question 3 for Grade Ten 375](#_Toc190951924)

[Table 10.A.10 Scale Score Distribution by Responses to Question 4 for Grade Ten 376](#_Toc190951925)

[Table 10.A.11 Scale Score Distribution by Responses to Question 1 for Grade Eleven 376](#_Toc190951926)

[Table 10.A.12 Scale Score Distribution by Responses to Question 2 for Grade Eleven 376](#_Toc190951927)

[Table 10.A.13 Scale Score Distribution by Responses to Question 3 for Grade Eleven 377](#_Toc190951928)

[Table 10.A.14 Scale Score Distribution by Responses to Question 4 for Grade Eleven 377](#_Toc190951929)

[Table 10.A.15 Scale Score Distribution by Responses to Question 1 for Grade Twelve 377](#_Toc190951930)

[Table 10.A.16 Scale Score Distribution by Responses to Question 2 for Grade Twelve 378](#_Toc190951931)

[Table 10.A.17 Scale Score Distribution by Responses to Question 3 for Grade Twelve 378](#_Toc190951932)

[Table 10.A.18 Scale Score Distribution by Responses to Question 1 for High School 378](#_Toc190951933)

[Table 10.A.19 Scale Score Distribution by Responses to Question 2 for High School 379](#_Toc190951934)

[Table 10.A.20 Scale Score Distribution by Responses to Question 3 for High School 379](#_Toc190951935)

[Table 10.A.21 Scale Score Distribution by Responses to Question 4 for Grades Ten and Eleven 379](#_Toc190951936)

[Table 10.A.22 Correlations Between Student Survey Responses and CAST Scale   
Scores 380](#_Toc190951937)

[Table 11.1 Item Development Results 383](#_Toc190951938)

### List of Figures

[Figure 4.B.1 PE distribution for Segment A of the CAST grade five assessment 49](#_Toc190780934)

[Figure 4.B.2 PE distribution for Segment A of the CAST grade eight assessment 50](#_Toc190780935)

[Figure 4.B.3 PE distribution for Segment A of the CAST high school assessment 51](#_Toc190780936)

[Figure 7.1 AI model-building and evaluation process 110](#_Toc190780937)

[Figure 7.2 Percentage of achievement levels 115](#_Toc190780938)

[Figure 8.H.1 Average CSEM by theta, grade five 316](#_Toc190780939)

[Figure 8.H.2 Average CSEM by theta, grade eight 320](#_Toc190780940)

[Figure 8.H.3 Average CSEM by theta, grade ten 324](#_Toc190780941)

[Figure 8.H.4 Average CSEM by theta, grade eleven 328](#_Toc190780942)

[Figure 8.H.5 Average CSEM by theta, grade twelve 332](#_Toc190780943)

[Figure 8.H.6 Average CSEM by theta, high school 336](#_Toc190780944)

Acronyms and Initialisms Used in the *California Science Test Technical Report*

|  |  |
| --- | --- |
| **Term** | **Definition** |
| AD | Assessment Development |
| ADEL | adult English learner |
| AERA | American Educational Research Association |
| AI | artificial intelligence |
| AIS | average item score |
| APA | American Psychological Association |
| ASL | American Sign Language |
| ATA | automated test assembly |
| CA NGSS | California Next Generation Science Standards |
| CA TAG | California Technical Advisory Group |
| CAA | California Alternate Assessment |
| CAASPP | California Assessment of Student Performance and Progress |
| CAI | Cambium Assessment, Inc. |
| CALPADS | California Longitudinal Pupil Achievement Data System |
| CalTAC | California Technical Assistance Center |
| CAST | California Science Test |
| CCC | crosscutting concept |
| *CCR* | *California Code of Regulations* |
| CDE | California Department of Education |
| CDS | county/district/school |
| CERS | California Educator Reporting System |
| CR | constructed response |
| CSEM | conditional standard error of measurement |
| DCI | disciplinary core idea |
| *DFA* | *Directions for Administration* |
| DIF | differential item functioning |
| DOK | depth of knowledge |
| DRM | data review meeting |
| *EC* | *Education Code* |
| ECD | evidence-centered design |
| EL | English learner |
| ELA | English language arts/literacy |
| ELPAC | English Language Proficiency Assessments for California |
| eSKM | Enterprise Score Key Management |
| ESS | Earth and Space Sciences |
| FIA | final item analysis |
| GPCM | generalized partial credit model |
| HOANR | highest obtainable adjusted number right |
| HONR | highest obtainable number right |
| IEP | individualized education program |
| IFEP | initial fluent English proficient |
| IMS | Instructional Management Systems |
| IRM | item review meeting |
| IRT | item response theory |
| ISAAP | Individual Student Assessment Accessibility Profile |
| K–12 | kindergarten through grade twelve |
| KSAs | knowledge, skills, and abilities |
| LEA | local educational agency |
| LOANR | lowest obtainable adjusted number right |
| LONR | lowest obtainable number right |
| LOSS | lowest obtainable scale score |
| LS | Life Sciences |
| MC | multiple choice |
| MCMS | multiple choice, multiple select |
| MCSS | multiple choice, single select |
| MH | Mantel-Haenszel |
| MSE | mean squared error |
| NCME | National Council on Measurement in Education |
| NGSS | Next Generation Science Standards |
| NR | number right |
| ONE | Online Network for Evaluation |
| OTI | Office of Testing Integrity |
| PAR | Psychometric Analysis & Research |
| PE | performance expectation |
| PPT | paper–pencil test |
| PRMSE | proportional reduction of mean squared error |
| PS | Physical Sciences |
| PT | performance task |
| QA | quality assurance |
| QTI | Question and Test Interoperability |
| QWK | quadratic-weighted kappa |
| RFEP | reclassified fluent English proficient |
| SBE | State Board of Education |
| SCOE | Sacramento County Office of Education |
| SD | standard deviation |
| SE | standard error |
| SEM | standard error of measurement |
| SEP | science and engineering practice |
| SFTP | secure file transfer protocol |
| SMD | standardized mean difference |
| SR | selected response |
| SRO | Scoring and Reporting Operations |
| SSID | Statewide Student Identifier |
| SSR | Student Score Report |
| STAIRS | Security and Test Administration Incident Reporting System |
| SVM | support vector machine |
| TDS | Test Delivery System |
| TEI | technology-enhanced item |
| TIF | test information function |
| TOMS | Test Operations Management System |
| TVIs | Teachers of the visually impaired |
| UAT | user acceptance testing |
| *USC* | *United States Code* |

**This page is intentionally left blank.**

## Introduction

This chapter provides an overview of the California Science Test (CAST) program, including background information, the purpose of the assessment, the intended population, and organizations and systems involved in CAST development and administration.

### Background

In October 2013, Assembly Bill 484 established the California Assessment of Student Performance and Progress (CAASPP) as the new student assessment system that replaced the Standardized Testing and Reporting program. The primary purpose of the CAASPP System of assessments is to assist teachers, administrators, and students and their parents/‌guardians by promoting high-quality teaching and learning using a variety of item types and assessment approaches. These assessments provide the foundation for the state’s school accountability system.

California adopted the California Next Generation Science Standards (CA NGSS) in September 2013. CAST is a computer-based assessment aligned with the CA NGSS. It was administered as a pilot for the first time during the 2016–17 CAASPP administration, followed by a field test administration during the 2017–18 CAASPP administration. The first operational CAST was administered during the 2018–19 CAASPP administration. The assessment is administered to students in grades five and eight and once in high school (that is, in grade ten, eleven, or twelve, if the student is not repeating grade twelve).

During the 2023–24 test administration, the CAASPP System comprised the following assessments:

* Smarter Balanced assessments and tools
* Summative Assessments—Computer-based assessments for English language arts/literacy (ELA) and mathematics in grades three through eight and grade eleven
* Interim Assessments—Optional resources developed for grades three through eight and grade eleven designed to inform and promote teaching and learning by providing information that can be used to monitor student progress toward mastery of the Common Core State Standards and that may be administered to students at any grade level
* Tools for Teachers—Professional development materials and instructional resources designed to help teachers use formative assessment processes for improved teaching and learning in all grade levels
* CAST and tools
* Summative Assessments—Computer-based assessments for science in grades five and eight and the high school grade band (that is, grade ten, eleven, or twelve, if the student is not repeating grade twelve)
* Interim Assessments—Optional resources developed for grades three through five, middle school, and high school that can be administered to any student at any grade level and are designed to measure smaller bundles of content to give teachers a better understanding of students’ science content knowledge and applicable skills and provide teachers with information about next steps for instruction
* Tools for Teachers—Instructional resources designed to help teachers use formative assessment processes for improved teaching and learning in all grade levels
* California Alternate Assessments (CAAs) for ELA and mathematics in grades three through eight and grade eleven for students with significant cognitive disabilities
* CAA for Science in grades five and eight and the high school grade band (that is, grade ten, eleven, or twelve, if the student is not repeating grade twelve) for students with significant cognitive disabilities
* The California Spanish Assessment, optional for eligible students in grades three through eight and the high school grade band and designed to measure a student’s literacy in Spanish language arts inclusive of reading, writing mechanics, and listening, as well as to serve as a high school measure suitable to be used in part for the California Seal of Biliteracy

CAST is presented as a computer-based assessment. Braille, large-print, and general paper–pencil test (PPT) versions of the CAST are made available to individual students within a local educational agency (LEA) whose need to take a PPT is documented in a student’s individualized education program or Section 504 plan or when a school experienced unexpected, temporary technology issues beyond the school’s control. Students who repeatedly experience difficulty accessing the computer-based assessments because of technical issues that cannot be resolved within two weeks may be allowed to take a standard PPT, upon approval by the California Department of Education (CDE).

More background information about the CAASPP System can be found on the CAASPP Description – *CalEdFacts* web page on the CDE website.

### Purpose of the Assessment

CAST assesses students with federally required science assessments in grades five and eight and once in high school. CAST is designed to assess the three dimensions (that is, science and engineering practices, disciplinary core ideas, and crosscutting concepts) of the CA NGSS by using various item types, some of which involve the use of dynamic stimuli and other types of new media (for example, animations of scientific phenomena, virtual engineering challenges, or simulated experiments).

The 2023–24 test administration was the fifth operational year of the assessment and the fifth year that Student Score Reports (SSRs) were generated (for more information, refer to [*Chapter 7: Scoring and Reporting*](#_Scoring_and_Reporting)).

### Assessment Design

The 2023–24 assessment, administered at each grade level or the high school grade band, was composed of three segments: A, B, and C. Content was assigned to students randomly, independent of their level of performance. Both discrete items and performance tasks (PTs) were included in the assessments.

At each grade level or within the high school grade band, the Test Delivery System (TDS) assigned students to two different operational item blocks in Segment A, with each block containing 13, 14, or 16 discrete items for grade five, grade eight, and high school, respectively. Each assessment also contained three different operational PTs in Segment B, representing each of the three CA NGSS domains—Earth and Space Sciences, Life Sciences, and Physical Sciences—with each task presenting five to six items. Finally, either one field test discrete item block (with six discrete items) or one field test PT (with six items) was assigned in Segment C.

The PTs were designed to provide students with an opportunity to demonstrate their ability to apply their skills in exploring and analyzing a complex, real-world scenario. The discrete items included traditional multiple-choice items, constructed-response (CR) items, and innovative technology-enhanced items. Refer to [*Chapter 3: Item Development and Review*](#_Item_Development_and) for details.

A braille form was available for students with visual impairments (refer to subsection[*4.6.1 Braille Form*](#_Braille_Form)). The braille form was composed of two Segment A blocks totaling 26, 28, or 32 discrete items for grade five, grade eight, and high school, respectively; and three Segment B PTs (totaling 17 PT items each for grade five, grade eight, and the high school grade band). The braille form had no field test items.

Table 4.1 in [chapter 4](#_Test_Assembly) lists the total number of unique items per segment across test forms.

### Intended Population

CAST was expected to be administered to approximately 1.5 million students in the general population. The intended population comprised all students in grades five and eight as well as one time for high school students in grade ten, eleven, or twelve who were assigned by their LEA (refer to subsection [*5.1.2 Student Test-Taking Requirements*](#_Student_Test-Taking_Requirements) for more details about the high school grade assignments).

Students eligible for alternate assessments take the CAA for Science in grades five and eight and once in the high school grade band. Analyses of the results of the CAA for Science are reported separately.

### Intended Use and Purpose of Test Scores

The results of assessments within the CAASPP System are used for two primary purposes as described in *Education Code* (*EC*) sections 60602.5(a) and (a)(4). (Excerpted from the *EC* Section 60602 web page.)

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

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

Therefore, the two primary purposes of an assessment within the CAASPP System are the following:

1. To communicate students’ progress in achieving the state’s academic standards to students, parents/guardians, and teachers
2. To inform decisions that teachers and administrators make about improving the educational program

Sections 60602.5(c) and (d) provide additional information regarding use and purpose of test scores for the system of assessments:

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

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

### Testing Window

CAST is administered within a testing window pursuant to *California Code of Regulations,* Title 5 (5 *CCR*)*,* Education, Division 1, Chapter 2, Subchapter 3.75, Article 2, sections 855(a)(1), 855(a)(2), 855(b), and 855(c). The CAST state testing window opened on January 9, 2024, and ended June 28, 2024.

Like other CAASPP assessments, CAST was untimed for students. A student could take the CAST within the LEA testing window over as many days as required to meet a student’s needs (5 *CCR*, Section 855[a][3]). The average time it took a student to complete and submit the assessment was approximately one hour and 57 minutes for grade five, one hour and 36 minutes for grade eight, and one hour and 10 minutes for high school.

### Significant CAST Developments in 2023–24

#### Updated Statewide Testing Window

5 *CCR* Section 855(a)(2) was amended to shift the end of the statewide testing window to June 30 (from July 15). For the 2023–24 test administration window, because June 30, 2024, occurred on a weekend, the end of the testing window was June 28, 2024.

#### Student Score Reports Redesign

Redesigned SSRs were made available for the 2023–24 test administration. Changes included the following:

* + - 1. SSR formats are PDF and HTML. For an HTML SSR, an LEA or parent or student portal vendor provided a link to a parent/guardian.
      2. Where applicable, results of a science assessment were included in the same PDF SSR as the results of the English language arts/literacy and mathematics assessments.
      3. All SSRs included comparisons to average student performance.

Additionally, Arabic was added as an available language.

#### Test Delivery

##### Changes to the Test Administrator Interface

The Test Administrator Interface was updated to a cleaner, more user-friendly appearance. This included a new functionality that allowed the test examiner to pin information for specific students to the top of the screen for monitoring.

##### Changes to Ending the Assessment in the Test Delivery System

The process for ending the assessment was streamlined. After the last question was presented, students selected [**Next**] (instead of [**End Test**]) to reach the review screen, which included the [**Submit Test**] button.

#### Accessibility Resources

The following accessibility resource–related updates were made:

* The definition of the non-embedded medical supports designated support was updated to allow Bluetooth hearing aids.
* The definition of the non-embedded amplification designated support was amended to remove noise buffers and white noise machines.

### Groups and Organizations Involved with the CAASPP System

#### California State Board of Education

The California State Board of Education (SBE) is the state agency that establishes educational policy for kindergarten through grade twelve in the areas of standards, instructional materials, assessment, and accountability. The SBE adopts textbooks for kindergarten through grade eight, adopts regulations to implement legislation, and has the authority to grant waivers of the *EC*.

In addition to adopting the rules and regulations for itself, its appointees, and California’s public schools, the SBE is also the state educational agency responsible for overseeing California’s compliance with programs that meet the requirements of the federal Every Student Succeeds Act that provides multiple measures of the academic performance and progress of schools on a variety of academic metrics (CDE, 2024c).

#### California Department of Education

The CDE oversees California’s public school system, which is responsible for the education of more than 5,800,000 children and young adults in more than 11,000 schools.[[1]](#footnote-2) California aims to provide a world-class education for all students, from early childhood to adulthood. The CDE serves the state by innovating and collaborating with educators, school staff, parents/guardians, and community partners which together, as a team, prepare students to live, work, and thrive in a highly connected world.

Within the CDE, it is the Instruction, Measurement, & Administration Branch that oversees programs promoting improved student achievement. Programs include oversight of statewide assessments and the collection and reporting of educational data (CDE, 2024b).

#### California Educators

A variety of California educators, including teachers and school administrators—who were selected on the basis of their qualifications, experiences, demographics, and geographic locations—were invited to participate in the various aspects of the assessment process. For the 2023–24 test administration, this included item review meetings, data review meetings, and scoring of the CAST CR items.

#### National Science Experts

ETS convenes a group of experts from various science backgrounds and disciplines on an as-needed basis. Together, the group discusses and provides insight on topics pertinent to CAST.

#### Contractors

A number of organizations contribute to the success of CAST.

##### Primary Testing Contractor—ETS

The CDE and the SBE contract with ETS to develop, administer, and report CAST. As the primary testing contractor, ETS has overall responsibility for working with the CDE to implement and maintain an effective assessment system and coordinating the ETS work with its subcontractors.

Activities conducted directly by ETS include, but are not limited to, the following:

* Providing management of the program activities
* Supporting and training county offices of education, LEAs, and direct funded charter schools
* Constructing, producing, and controlling the quality of PPT booklets and related test materials
* Developing processes and scripts associated with remote testing
* Hosting and maintaining a website with resources for LEA CAASPP coordinators
* Developing, hosting, and providing support for the Test Operations Management System (TOMS)
* Supporting the California Educator Reporting System (CERS)
* Processing student test assignments
* Processing orders and shipment of test materials
* Servicing all aspects of CR scoring for CAST
* Producing and distributing score reports electronically
* Developing a summary score reporting website that can be viewed by the public
* Completing all psychometric procedures
* Providing a tiered help desk support system for LEAs

##### Subcontractor—Cambium Assessment, Inc.

ETS also monitors and manages the work of Cambium Assessment, Inc. (CAI), subcontractor to ETS for the CAASPP System of computer-based assessments. Activities conducted by CAI include

* providing the CAI proprietary TDS, including the Student Testing Interface, Test Administrator Interface, secure browser, and practice and training tests;
* hosting and providing support for its TDS, a component of the overall CAASPP Assessment Delivery System;
* hosting and providing support for the Data Entry Interface, the web browser–based application that, for the operational administration of the CAST, allows users to enter student responses;
* scoring machine-scorable items; and
* providing high-level technology help desk support to LEAs for technology issues directly related to the TDS.

##### Subcontractor—Sacramento County Office of Education

ETS contracted with the Sacramento County Office of Education to manage all activities associated with educator recruitment, training, and outreach, including the following:

* Supporting and training county offices of education, LEAs, and charter schools
* Developing informational materials
* Recruiting and providing logistics for educator meetings
* Producing test administration scripts

### Systems Overview and Functionality

#### Test Operations Management System

TOMS is the password-protected, web-based system used by LEAs to manage all aspects of CAASPP testing. TOMS serves various functions, including, but not limited to, the following:

* Managing test administration windows
* Assigning and managing CAASPP online user roles
* Managing student test assignments and accessibility resources
* Ordering test materials
* Viewing and downloading reports
* Reporting security incidents
* Providing a platform for authorized user access to secure materials, such as CAASPP *DFAs*, student data and results, CAASPP user information, and access to the CAASPP Security and Test Administration Incident Reporting System/Appeals process

TOMS receives student enrollment data and LEA and school hierarchy data from the California Longitudinal Pupil Achievement Data System (CALPADS) via daily feed. CALPADS is “a longitudinal data system used to maintain individual-level data including student demographics, course data, discipline, assessments, staff assignments, and other data for state and federal reporting.”[[2]](#footnote-3)

LEA staff involved in the administration of the CAASPP—such as LEA CAASPP coordinators, site CAASPP coordinators, test administrators, and test examiners—are assigned varying levels of access to TOMS. For example, only an LEA CAASPP coordinator is given permission to assign and manage user roles; a test administrator or test examiner cannot download student reports. A description of user roles is explained more extensively in the *2023–24 CAASPP Online Test Administration Manual* (CDE, 2024a).

#### Test Delivery System

The TDS is the means by which the statewide computer-based assessments are delivered to students. Components of the TDS include

* the Test Administrator Interface, the web browser–based application that allows test administrators to activate student assessments and monitor student testing;
* the Student Testing Interface, on which students take the assessment using the secure browser; and
* the secure browser, the computer-based application through which the Student Testing Interface may be accessed. (The secure browser prevents students from accessing other applications during testing.)

#### Practice and Training Tests

All California testing programs have practice and training tests to inform educators, parents/‌guardians, and students about the individual assessments. The practice and training tests were provided to LEAs to prepare students and LEA staff for administration of the CAST. These tests simulated the experience of CAST computer-based assessments. Unlike the summative assessments, the practice and training tests did not gauge student success on the operational assessment, or produce scores. Students, teachers, and the public could access the practice tests and training tests using a web browser, although accessing them through the secure browser permitted students to test with different embedded accommodations, such as text-to-speech; and to try out different assistive technology. When remote testing was added as an optional means of test administration, the practice and training tests permitted test administrators and students to practice using the remote monitoring and communication features.

The purpose of the training tests is to allow students and test administrators to quickly become familiar with interacting with the user interface, different item types, and components of the TDS as well as with the process of starting and completing a testing session.

The purpose of the practice tests is to allow students and test administrators to experience a grade-level assessment, grade-specific items, and difficulty levels; and become familiar with the format and structure of an operational assessment.

A purpose of both the practice and training tests is to provide an opportunity for educators to assign embedded designated supports and accommodations and determine how they worked for their students prior to using the resources in an operational test setting.

#### California Educator Reporting System

CERS is the system used by LEAs to view student results from CAASPP testing as they became available. The primary purpose of CERS is to provide educators and administrators with access to timely assessment results for individual students and groups of students.

CERS allows educators to view their students’ test results at the individual student level and at the aggregate level using grouping and other features. For example, educators can create customized groups from assigned student groups based on student demographic information or other characteristics of their choosing. The student results sent to CERS are appropriate for analysis of assessment results for use in informing instruction and local planning.

#### Test Results for California’s Assessments Website

The Test Results for California’s Assessments website is used by educators, families, researchers, and interested members of the public to view aggregate results from the CAST. The primary purpose of the Test Results for California’s Assessments website is to provide users with access to results data for groups of students and to allow comparison of test result data for various student groups. Test scores for a given grade level are aggregated at the school, LEA or direct funded charter school, county, and state levels. The aggregate scores are generated for selected student groups of interest (for example, gender, ethnicity, economic status, migrant status, and disability status) and for the total population.

#### Constructed-Response Scoring Systems for ETS

CR items from the TDS were routed to ETS CR scoring systems. CR items were scored by certified raters or the artificial intelligence (AI) scoring engine. More information regarding scoring of CR items is available in [*Chapter 7: Scoring and Reporting*](#_Scoring_and_Reporting).

For CAST, targeted efforts were made to hire qualified raters from existing CAASPP rater pools and California science teachers. The hired human raters were provided with in-‍depth training and were certified before starting the scoring process. Human raters were organized under a scoring leader and were provided CAST scoring materials such as benchmark sets, training sets, scoring rubrics, and scoring notes. The quality-control processes for CR scoring are explained further in [*9.5 Quality Control of Scoring*](#_Quality_Control_of_2).

The CR items could also be rated by AI scoring engines (for example, the ETS *c-rater*™ system). The *c-rater*™ system uses state-of-the-art, machine-learning technology to score items that elicit and measure knowledge about specific content. The use of such engines requires models to be built with reliable human-rating data. For the 2017–18 test administration, a data collection design was used to provide data to build and evaluate AI models for the field test CRs. For details on AI model building and evaluation on field test CRs, refer to the *California Science Test Field Test Technical Report 2017–2018 Administration* (CDE, 2019).

During the first operational administration in 2018–19, AI scoring was used to score responses for those CRs with approved AI models. A careful data collection design was also used to provide data to build the AI scoring engine for future use. Additional scoring models were built in 2019, 2021, 2022, 2023, and 2024 for use in future test administrations. Note that no new AI scoring models were built in 2020 because the 2019–20 CAST administration was suspended because of the novel coronavirus disease 2019 pandemic, so no range finding activities were conducted that year.

A CR sampling plan to support the new AI model building is provided in subsection [*7.1.1.1.2 Sampling Process for Field Test Constructed-Response Items*](#_Sampling_Process_for). An additional 37 CAST models for automated scoring were built in 2024 that were based on operational data and 2023–24 field test data. Refer to subsection [*7.1.1.3 Artificial Intelligence Model Building*](#_Artificial_Intelligence_Model) for more information.

The engine computes a large set of linguistic features from each response that relate to the content focus of the item. This broad set of features extends beyond key words to capture grammatical relationships and mitigates the impact of spelling and grammatical variation on how the model assigns scores. The *c-rater*™ system evaluates expected content for subject-matter CR items in content areas, including social studies, science, ELA, and mathematics.

The ETS process required test designers to define the required content but did not ask them to predict every aspect of the form of student language. The *c-rater*™ engine filtered out potential, nonscorable responses (for example, responses in a language other than English, no-‍attempt responses such as “I don’t know,” etc.). Filtering was applied both during the AI-‍scoring and the model-building steps to ensure AI-scoring models were built on reliable data; and when the AI-scoring model was deployed, to ensure that such responses were filtered and scored correctly.

Any response that was entirely in a language other than English as detected by *c-rater*™ was given a specific advisory designation and handled following the policy established with the CDE to mark these responses as not scorable and return them to the Online Network for Evaluation with an advisory code to be human-scored. If the response was in Spanish, these responses were then reviewed by Spanish biliterate raters and scored according to the rubric. If the response was not in English or Spanish, the response received a zero score.

### Overview of the Technical Report

This technical report addresses the characteristics of the CAST administered in spring 2024 and contains 10 additional chapters as follows:

* [Chapter 2](#_Overview_of_CAST_1) presents an overview of processes involved in a CAST testing cycle. This includes item development, test assembly, test administration, fairness and accessibility, generation of test scores, and psychometric analyses.
* [Chapter 3](#_Item_Development_and) discusses the detailed procedures of item development and review for CAST to help ensure valid interpretation of test scores.
* [Chapter 4](#_Test_Assembly) discusses the content, psychometric criteria, and reviews that guide procedures of CAST test assembly.
* [Chapter 5](#_Test_Administration) details the processes involved in the administration of CAST. It also describes the procedures followed by ETS to maintain test security throughout the test administration process.
* [Chapter 6](#_Standard_Setting) presents a high-level overview of the standard setting procedures implemented for CAST.
* [Chapter 7](#_Scoring_and_Reporting) summarizes the types of scores and score reports that are produced at the end of each administration of CAST.
* [Chapter 8](#_Psychometric_Analyses) summarizes the statistical procedures and results for 2023–24. These analyses include

test-taking rates,

classical item analyses,

differential item functioning analyses,

item response theory analyses,

response time analyses,

reliability analyses, and

validity analyses.

* [Chapter 9](#_Quality-Control_Procedures) highlights the quality-control processes used at various stages of development and administration of CAST.
* [Chapter 10](#_Student_Survey_1) describes the development and administration of the survey questionnaires for students and the results of analyses of their responses.
* [Chapter 11](#_Continuous_and_Systematic) discusses the various procedures used to gather information to improve CAST as well as strategies to implement possible improvements.

### References

*California* *Code of Regulations*,Title 5,Education, Division 1, Chapter 2, Subchapter 3.75, Article 2, Section 855.

California Department of Education. (2019). *California Science Test field test technical report 2017–18 administration* [Unpublished report]. Sacramento, CA: California Department of Education.

California Department of Education. (2024a). *CAASPP online test administration manual*. Sacramento, CA: California Department of Education.

California Department of Education. (2024b, August). *Organization*. California Department of Education website.

California Department of Education. (2024c, October). *State Board of Education responsibilities*. California Department of Education website.

California *Education Code*, Title 2, Elementary and Secondary Education, Division 4, Part 33, Chapter 5, sections 60602.5(1) and 60602.5(a)(4). (n.d.).

Every Student Succeeds Act of 2015, 20 U.S.C. 20 § 6301 (2015).

## Overview of CAST Processes

This chapter provides an overview of the processes implemented by ETS during a typical, full testing cycle for the California Science Test (CAST), including item development, test design, test administration, and scoring. The details on each step in the process will be presented in the subsequent chapters.

### Item Development

CAST item development processes sustained best practices from prior development cycles and incorporated further innovations informed by feedback from the California Department of Education (CDE) and educators in the field. For CAST, items and associated stimuli were developed to integrate the dimensions of the performance expectations (PEs) while maintaining grade-level appropriateness for test takers. Item review meetings with California educators were instrumental in determining both the proper integration of the PE dimensions and grade-level appropriateness.

This section describes the process used to develop new items and how California educators are selected to participate in the process during typical item development cycles.

#### Design Guidelines

ETS content specialists refer to design patterns and task templates as part of the emerging evidence-centered design documentation created by ETS researchers and based on current educational research to properly frame the construct measured in each item (Mislevy, Almond, & Lukas, 2003). As such, all items used in the 2023–24 CAST administration were appropriate for the grade level and aligned with the California Next Generation Science Standards (CA NGSS).

#### Content Guidelines

Throughout the item writing process, ETS developers adhere to the ETS foundational guidelines for quality item writing. These guidelines form the basis for training item writers and for the rigorous review process that is implemented for every item. Additionally, item content specifications and the CA NGSS PEs are used to guide the writing of items for CAST. Refer to section [*3.2 Guidelines*](#_Guidelines) for the guidelines of item writing, including the item content specifications.

ETS trains California science teachers to develop items for CAST. California science teachers are instructed to produce items that span a variety of science and engineering practices and science domains (that is, Earth and Space Sciences; Life Sciences; Physical Sciences; and Engineering, Technology, and Applications of Science) to provide as wide an array of items as possible for CAST forms construction. A key factor in determining the assignment of PEs to each item writer was the teaching experience and expertise that the item writer possessed.

#### Item Types Guidelines

CAST was designed to assess the CA NGSS using discrete items, single and multipoint items, and performance tasks (PTs). A variety of item types were developed, including traditional multiple-choice (MC) items, constructed-response (CR) items, some familiar technology-enhanced item (TEI) types, as well as PT stimuli that used simulations and animations. Refer to section [*3.2 Guidelines*](#_Guidelines) for more details on the number of items developed, and to subsection [*3.1.4 Item Types and Features*](#_Item_Types_and) for the types of items used in CAST. A significant emphasis was placed on filling the CAST item bank with items that have students explore phenomena using item types that best fit the construct. ETS also generated item sets—PTs—internally to measure more complex skills in a particular domain.

### Test Assembly

The 2023–24 CAST design was based on the California State Board of Education–approved, high-level test design for an operational assessment, which requires that all students in the tested grade levels or grade band participate in three segments of the assessment: Segment A, Segment B, and Segment C. The first two segments made up the operational assessment; Segment C was for field-testing future operational items. Note that braille computer-based forms and paper–pencil tests did not include Segment C.

ETS designed the general CAST forms to be taken in approximately two hours and used historical timing data from previous CAST assessments that had the same item types to estimate the amount of time needed to complete MC, CR, and TEI types. [*Chapter 4: Test Assembly*](#_Test_Assembly) provides details about test assembly.

#### Test Blueprint

Blueprints represent a set of constraints and specifications to which each test form must conform. CAST had three main subcontent areas or domains: Earth and Space Sciences, Life Sciences, and Physical Sciences. The blueprint for the assessment is shown in [appendix 4.A](#_Appendix_4.A:_Test_1), table 4.A.1 through table 4.A.6.

For the 2023–24 assessment, each grade five student took 8 or 9 discrete items from each domain (blueprint allows eight to nine items), for a total of 26 items that were worth 31 total points; and three PTs assessing different domains, worth 20 to 21 total points. Each grade eight student took 9 or 10 discrete items from each domain (where the blueprint allows 8 to 10 items), for a total of 28 items that were worth 32 total points; and three PTs assessing different domains, worth 19 to 20 total points. Finally, each high school student took 10 or 11 discrete items from each domain (blueprint allows 9 to 12 items), for a total of 32 items that were worth 37 total points; and three PTs assessing different domains, worth 20 to 21 total points.

Each student taking the general, computer-based form during the 2023–24 test administration received a field test block of six discrete items or a field test PT with six items.

### Test Administration

CAST was administered using the secure browser and Test Delivery System (TDS), ensuring a secure, confidential, standardized, consistent, and appropriate administration for students. Additional information about the administration of CAST can be found in [*Chapter 5: Test Administration*](#_Test_Administration).

Testing could occur in person and remotely. Students receiving in-person instruction were tested in person, at a school site. Remote administration, which is intended as an option for a local educational agency (LEA) only when its students are receiving remote instruction, occurred when either the students, test administrator, or both were located at different physical locations. In remote testing, the test administrator monitors students’ progress throughout the assessment by using remote monitoring tools connected to the TDS.

#### Test Security and Confidentiality

All operational assessments within the California Assessment of Student Performance and Progress (CAASPP) System are secure. For CAST administration, every person having access to test materials maintained the security and confidentiality of the assessments. The ETS internal Code of Ethics requires that all test information, including tangible materials (such as test booklets, test items, and test results), confidential files, processes, and activities were kept secure. To ensure security for all assessments 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 subsection[*5.6.1 The ETS Office of Testing Integrity*](#_The_ETS_Office_1) in [*Chapter 5: Test Administration*](#_Test_Administration).

In the pursuit of enforcing secure practices, ETS strives to safeguard the various processes involved in an assessment development and administration cycle. Those processes are listed next. The practices related to each of the following security processes are discussed in detail in section [*5.6 Test Security and Confidentiality*](#_Test_Security_and):

* Procedures to maintain standardization of test security
* Test security monitoring
* Security of electronic files using a firewall
* Transfer of scores via secure data exchange
* Data management in the secure database
* Statistical analysis on secure servers
* Student confidentiality
* Student test results

#### Procedures to Maintain Standardization

ETS takes all necessary measures to ensure the standardization of administration of CAST.

CAST is administered in conjunction with the other assessments that compose the CAASPP System. ETS employs processes to ensure the standardization of an administration cycle; these processes are discussed in more detail in section [*5.2 User Roles and Standardization*](#_User_Roles_and).

Staff at LEAs involved in the CAASPP administration include LEA CAASPP coordinators, site CAASPP coordinators, and test administrators. The responsibilities of each of the staff members are described in the *Test Operations Management System* *(TOMS) User Guide* (CDE, 2024b).

Several series of instructions regarding the CAASPP administration are compiled in detailed manuals and provided to the LEA staff. Such documents include, but are not limited to, the following:

* ***CAASPP Online Test Administration Manual*—**This web-based manual provides test administration procedures and guidelines for LEA CAASPP coordinators and site CAASPP coordinators, as well as the script and *Directions for Administration* (*DFA*) to be followed exactly by test administrators during a testing session (CDE, 2024a). (Refer to [*5.2.4.2 CAASPP Online Test Administration Manual*](#_CAASPP_Online_Test) in [chapter 5](#_Test_Administration) for more information.)
* ***TOMS User Guide*—**This web-based manual provides instructions for TOMS, allowing LEA staff, including LEA CAASPP coordinators and site CAASPP coordinators, to perform several tasks, including setting up test administrations, adding and managing users, assigning assessments, and configuring computer-based student test settings (CDE, 2024b). (Refer to [*5.2.4.3 Test Operations Management System User Guide*](#_Test_Operations_Management_1) in [chapter 5](#_Test_Administration) for more information.)

### Fairness and Accessibility

Several procedures are in place to ensure that the CAST is fair and accessible to all students. This section provides information on the available accessibility resources.

#### Overview

All eligible students enrolled in a California public school participate in the CAASPP System of assessments, including students with disabilities and English learner (EL) students. Additional resources are sometimes needed for these students. The CDE provides a full range of assessment resources for all students, including those who are EL students and students with disabilities.

#### Student Accessibility Resources

There are four different categories of student accessibility resources in the California assessment accessibility system, including universal tools, designated supports, accommodations, and unlisted resources that are permitted for use in CAASPP computer-based assessments. These are listed in the CDE California Assessment Accessibility Resources Matrix (Accessibility Matrix) (CDE, 2023).

**Universal tools** are available to all students. These resources may be turned on and off when embedded as part of the technology platform for the computer-based CAASPP on the basis of student preference and selection.

**Designated supports** are available to all students when determined as needed by an educator or team of educators, with parent/guardian and student input as appropriate, or when specified in the student’s individualized education program (IEP) or Section 504 plan.

**Accommodations** must be permitted on the CAASPP for all eligible students when specified in the student’s IEP or Section 504 plan.

**Unlisted resources** are non-embedded and made available if specified in the eligible student’s IEP or Section 504 plan and do not jeopardize test security, and only on approval by the CDE. An unlisted resource may change the construct being measured.

[Appendix 5.A](#_Appendix_5.A:_Accessibility_1) presents counts and percentages of students assigned designated supports, accommodations, and unlisted resources for the 2023–24 CAST administration. The tables in [appendix 5.A](#_Appendix_5.A:_Accessibility_1) were created using student demographic data in version 3 of the production data file (“P3”) updated on October 7, 2024.

The majority of students did not use any designated supports, accommodations, or unlisted resources.

#### Description of Differential Item Functioning

Differential item functioning (DIF) analyses are conducted to detect possible test bias by locating items for which one group of students performs significantly better than another group. DIF is a collection of statistical methods used to recognize whether performance varies across different groups of students (for example, male versus female or White versus Black or African American). If an item performed differentially across student groups, even when students were matched on ability, the item may be measuring something other than the intended construct. Therefore, it is important to identify items flagged for DIF. Content experts and bias and sensitivity experts from diverse backgrounds reviewed these DIF-flagged items to determine the potential sources and meanings of performance differences. Refer to section [*8.3 Differential Item Functioning Analyses*](#_Differential_Item_Functioning) for additional information about DIF.

### Scores

Individual student scores were reported for the 2023–24 CAST administration. Student performance on the reporting scale was designated into one of the four achievement levelsdescribed in subsection [*7.3.1 Total-Test Achievement Levels*](#_Total-Test_Achievement_Levels). For information regarding score specifications and score reports, refer to [*Chapter 7: Scoring and Reporting*](#_Scoring_and_Reporting).

#### Score Reporting

TOMS is a secure website hosted by ETS that permits LEA users to manage aspects of CAASPP test administration such as test assignment and the assignment of test settings. TOMS also provides a secure means for LEA CAASPP coordinators to download Student Score Reports as PDF files.

CAST scores can also be viewed through the California Educator Reporting System (CERS), a secure website that provides authorized users with interactive and cumulative online reports for CAST at the student, school, and LEA levels. CERS also provides individual score reports. Refer to subsection [*7.4.1 Online Reporting*](#_Online_Reporting_1) for details about TOMS and CERS and subsection [*7.4.3 Types of Score Reports*](#_Types_of_Score) for the content of each type of score report.

#### Aggregation Procedures

To provide meaningful results to interested educators, CAST scores for a given grade-level—or in some cases, grade-band— assessment were aggregated at the school, LEA or direct funded charter school, county, and state levels. State-level results are available on the Test Results for California’s Assessments website. The aggregate scores were presented for all students or selected demographic student groups.

Aggregate scores were generated by combining student scores at the state, LEA or direct funded charter school, or school level; combining student scores for all students; or by combining student scores for students who represent selected demographic student groups.

The aggregation procedures used to present CAST results are described in section [*7.3 Summary Statistics*](#_Summary_Statistics). Aggregate results by demographic variables are presented in [appendix 7.D](#_Appendix_7.D:_Demographic). In table 7.D.1 through table 7.D.6, students are reported by demographic groups, including gender, ethnicity, English language fluency, disability status, and economic status, as well as crosstab analysis for ethnicity and economic status. The tables show the numbers of students with valid scores in each group, scale score means and standard deviations, and the percentage of students in each achievement level. To protect student privacy, statistics are presented in the tables as “N/A” when the number of students in the sample is 10 or fewer. Definitions for the demographic student groups included in these tables are provided in table 7.17.

### Psychometric Analyses

Psychometric analyses were conducted on the data from CAST, including classical item analyses, DIF analyses, item response theory (IRT) calibration and linking, testing time analyses, and reliability analyses. The results of these analyses support understanding of item performance and internal structure of the assessment and provide validity evidence for both response processes and scoring. Detailed descriptions of these analyses are presented in [*Chapter 8: Psychometric Analyses*](#_Psychometric_Analyses).

#### Description of Classical Item Analyses

The psychometric analyses for the CAST data included classical item analyses and DIF analyses to evaluate the performance of the operational items and the embedded field test items. The classical item analyses included the computation of item difficulty indices, the item-total correlation indices, the omission rate of each item, and the proportion of students obtaining each score point for polytomous items. CDE-approved flagging rules based on these statistics identified items that were not performing as expected. A description of the classical item analyses procedure is provided in section [*8.2 Classical Item Analyses*](#_Classical_Item_Analyses)*.* A description of the DIF analyses procedure is provided in section [*8.3 Differential Item Functioning Analyses*](#_Differential_Item_Functioning)*.*

#### Description of Item Response Theory Analyses

IRT is used to calibrate items, link item parameter estimates, scale or equate test scores across different forms or test administrations, evaluate item performance, build an item bank, and assemble test forms. Detailed information on the models and the procedures for the calibration and linking analyses are included in section [*8.4 Item Response Theory Analyses*](#_Item_Response_Theory).

### References

California Department of Education. (2023). *California assessment accessibility resources matrix*. California Department of Education website.

California Department of Education. (2024a). *CAASPP online test administration manual*. Sacramento, CA: California Department of Education.

California Department of Education. (2024b). *Test Operations Management System user guide*. Sacramento, CA: California Department of Education.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* (ETS Research Report RR-03-16). Princeton, NJ: Educational Testing Service.

## Item Development and Review

This chapter discusses the detailed procedures of item development for the 2023–24 California Science Test (CAST) administration.

### Use of Evidence-Centered Design

The principles and practices of evidence-centered design (ECD) guided the development of all CAST items. Developed at ETS in 1999, ECD is a framework for designing, producing, and delivering educational assessments so that evidence collected about student performance during testing provides support for claims about what students know and can do.

#### Principles

ECD is an important tool used to support assessment validity arguments as well as inferences made about student scores (Mislevy, Almond, & Lukas, 2003).

As described in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), a coherent validity argument, including alignment evidence, is essential to supporting the appropriateness of inferences made on the basis of an assessment’s results. ETS consulted with California educators, the Stanford Next Generation Science Standards (NGSS) Assessment Project, and other national leaders of science education reform in building CAST development processes. By employing ECD during the development process, ETS built the validity argument needed to support the operational use of CAST.

#### Theory of Action Model

One of the priorities of interested educators presented to the California State Board of Education (SBE) in March 2016 was the “focus on providing information to support the *improvement of teaching and learning*” (California Department of Education [CDE], 2016).

The principles of ECD pervade all aspects of CAST, including item development, so that CAST is able to gather evidence of student proficiency that can be used to support improvement in how science is taught. Because CAST items are aligned with a wide variety of performance expectations (PEs), CAST allows students to demonstrate a wide array of scientific knowledge and skills. In this way, CAST supports instruction that “encourages students to build the knowledge and skills needed for college and careers” (CDE, 2019).

As part of ECD, ETS continually analyzes student performance data and feedback from interested California educators to improve item alignment with the standards and provide students with grade-level appropriate phenomena. Additionally, ETS incorporates feedback from interested California educators in the CAST development process to allow for continual improvement of the assessment and to impact instructional strategies. These processes support the validity argument for CAST and support the claims enumerated in the CAST blueprint.

The logic model provides the sequence of how CAST was conceptualized, starting with the components that led to the design and development of the assessment, the anticipated actions by interested educators, and outcomes (both intended for the intermediate and long-term futures; and potentially unintended).

##### Components

CAST was aligned with the California Next Generation Science Standards (CA NGSS) to assess rigorous standards that emphasize continual building of knowledge and skills, as well as to

* assess both the breadth and depth of the CA NGSS via items aligned with the three dimensions (that is, science and engineering practices [SEPs], disciplinary core ideas [DCIs], and crosscutting concepts [CCCs]) of the standards; and
* include performance tasks (PTs) that encourage test takers to thoroughly explore a phenomenon through the dimensions of the CA NGSS, in addition to discrete items.

Accessibility resources are available to students to ensure equity of the assessment. Refer to section [*5.4 Accessibility Resources*](#_Accessibility_Resources_1) for a description of the resources that were available to students taking the CAST.

##### Actions

***Interested participants*** are an essential part of ECD. Interested participant groups involved in the logic model are students, educators, and those developing the assessment.

***Students*** engage with PTs and discrete items that provide the opportunity to experience authentic science.

***Educators*** will

* participate in item writer workshops, item review meetings (IRMs), data review meetings (DRMs), and range finding activities to improve understanding of the multidimensionality of the CA NGSS;
* access item content specifications to develop aligned classroom assessments; and
* access released practice and training tests to better understand the item types and range of science content assessed on CAST.

***Assessment developers*** refine development practices and ways to obtain and apply external feedback after reviewing field test item performance data.

##### Intermediate Outcomes

After test administration, the results from CAST can show how students have begun to make sense of phenomena using the knowledge and skills learned through instruction aligned with the CA NGSS.

Results from CAST

* provide educators, students, and parents/guardians with information about the student’s progress; and
* help advise schools and local educational agencies (LEAs) on strengths and weaknesses in their instructional programs to provide better alignment with the CA NGSS.

##### Long-Term Outcomes

Both students and interested educator groups can benefit from a science assessment developed using ECD.

***Students*** develop the ability to provide mechanistic reasoning about phenomena in the natural and designed world around them.

***Educators*** continue to better align instruction with the three dimensions of the CA NGSS to promote greater science proficiency.

##### Unintended Outcomes

As with any endeavor, there may be unintended outcomes.

For example, in early administrations of CAST, teachers and students will have had limited access to teaching and learning of the CA NGSS. Therefore, score users may have initially misinterpreted CAST results because of limited access to the CA NGSS.

This will be resolved as student exposure to the CA NGSS increases and educators have put effective strategies in place for presenting the concepts associated with the CA NGSS. As described in subsection [*8.7.6 Consequences of Testing*](#_Consequences_of_Testing), potential unintended outcomes included reduced morale among teachers and students and increased pressure on students, which could lead to higher school dropout rates.

#### Incorporation into Item Development Processes

For the CAST item development process, ETS began with the existing NGSS evidence statements that provide additional detail on what students should know and be able to do and that describe the NGSS PEs in some detail (Achieve, 2015). ETS drafted work on the task models and task templates to outline the types of items that would elicit student output sufficient to provide observable evidence of achievement in each assessed PE.

The task-model documentation is practice based. ETS developed one design pattern for each CA NGSS SEP and began developing one to three task templates for each design pattern. Each design pattern captured the results of domain analysis by specifying knowledge, skills, and abilities (KSAs) focal to the corresponding SEP, characteristics of the SEP that differ across the two grade levels and the high school grade band (that is, grade ten, eleven, or twelve), and characteristic features of assessments that elicit evidence of the focal KSAs.

During the drafting stage, ETS further specified approaches to the task templates designed to engage students meaningfully with the SEP by specifying item characteristics, work products, and observations that can be made about student proficiency from those work products. The approaches were used during both item development and revision to ensure that the student responses elicited by the items validly reflected the integrated science understanding specified in the targeted PEs. Detailed information on item content specifications is presented in subsection [*3.2.3 Item Specifications*](#_Item_Specifications).

ECD is an inherently iterative process. Lessons learned in one stage are used to refine both test design decisions and documentation for later stages. Information documented in some artifacts that were key to the development of the CAST items was later incorporated into more comprehensive documents. For example, the information contained in the design patterns described previously was, for later rounds of item development, incorporated into more robust item content specifications. Item content specifications for each PE assessed on CAST include assessment targets, framed from focal KSAs, for each dimension of the PE.

Similarly, the definition of claims for CAST is an ongoing and iterative process, one informed both by the data collected from the CAST field test administration and the data collection from operational administrations. Comprehensive documentation of this process is captured in a white paper titled *Use of Evidence-Centered Design in CAST Item and Test Development* (ETS, 2019b).

#### Item Types and Features

Every CAST item assessed a CA NGSS DCI as well as at least one of the other two CA NGSS dimensions (that is, SEP or CCC). Wherever possible, a single item assessed all three dimensions of the CA NGSS. However, leading NGSS experts agreed that it was not always practical to assess all three dimensions using a single item (ETS, 2016b).

ETS used item types, individually and in combinations or sets, to measure targeted CA NGSS content. In some cases, the presentation of the content involved the use of dynamic stimuli and other types of media (for example, animations of scientific phenomena, real-life engineering challenges, and simulated experiments run multiple times by a student to generate data for analysis) to provide rich opportunities for students to demonstrate their scientific knowledge and skills.

For the item development process, ETS developed item types and features for the 2023–24 CAST that were supported by *Instructional Management Systems (IMS) Global Question and Test Interoperability (QTI)* standards (IMS, 2020).

Table 3.1 outlines the major categories of QTI item types that were included in CAST. This includes item types ranging from traditional multiple-choice (MC) items and constructed-response (CR) items (that is, extended text) to technology-enhanced item (TEI) types (the remainder of the item types).

Table 3.1 Item Types Used in CAST

|  |  |
| --- | --- |
| **Feature** | **Description** |
| **Multiple Choice, Single Select (MCSS)** | The MCSS item type generally consists of a stem and list of choices; the student can select only one choice (option) to respond. The options have radio buttons, but the student can select text or an image by clicking in the option space rather than exactly on the radio button. Items should have four options: one key and three distractors. |
| **Multiple Choice, Multiple Select (MCMS)** | The MCMS item type generally consists of a stimulus, stem, and a list of choices; the student can select one or more choices (options) to respond. The options use checkboxes, but the student can select text or an image by clicking in the option space rather than exactly in the checkboxes. The stem should specify how many selections should be made. An item with a two-option keyset should have four or five total options, and an item with a three-option keyset should have five options. |
| **CR** | The CR item type consists of a stem to which the student must provide a typed response in a designated response box. |
| **Composite** | The composite item type is a multipart item that is expected to have two responses selected from two sets of options. This item type contains two item parts from the machine-scored list: one MC item part and one TEI part. The TEI part should abide by guidelines provided in this document for grid and inline choice list items. |
| **Grid, Multiple Select** | The student responds to the grid, multiple-select item type by placing a check mark in two or more cells in a grid. Grid items generally should have three rows and three columns and should not have more than four rows and four columns or 16 total checkboxes. |
| **Inline Choice List, Single Select** | In the inline choice list, single-select item type, the stem contains a single blank that the student must fill in by selecting a single choice from a drop-down list. |
| **Inline Choice List, Multiple Select** | In the inline choice list, multiple-select item type, the stem contains two or more blanks that the student must fill in by selecting choices from a drop-down list of choices. |
| **Interactive** | In the interactive item type, unscored simulations require the student to select inputs that provide certain outputs to respond to one or more items and scored simulations require the student to select inputs that are scored. |
| **Zone, Single Select** | The zone, single-select item type has answer choices that are predefined “hot spots” or zones on an image. When the student clicks on a hot spot, the selection is shaded light blue with a blue outline. The student selects one zone to respond. *This item type is no longer under development.* |
| **Match, Multiple Select** | In the match, multiple-select item type, the student responds by dragging and dropping two or more choices (sources) into the appropriate locations (targets). There should be a maximum of five sources to fill a maximum of four targets *This item type is no longer under development.* |
| **Bar-Picturegraph, Multiple Select** | In the bar-picturegraph, multiple-select item type, the student responds by dragging the top line of three to five bars in a bar graph or histogram to specific heights. *This item type is no longer under development.* |

### Guidelines

Each item for CAST was developed through a comprehensive development cycle and designed to conform to principles of item writing defined by ETS and approved by the CDE. Each item in the CAST operational item bank was developed to measure a specific CA NGSS PE. In addition, guidelines for style, fairness, and bias and sensitivity helped item developers and reviewers to ensure consistency across the item development process.

#### Plan

The item development plan for CAST focused on developing items that integrated the DCI and at least one of the other two dimensions of the CA NGSS—SEPs and CCCs. The plan incorporated a diverse selection of PEs to incorporate a range of SEPs, DCIs, and CCCs.

Certain PEs were the focus of discrete item development, with the ultimate goal of generating operational items to enrich the CAST item bank and meet the PE coverage goals stated in the blueprint. For the most part, each PT was developed to assess PEs that did not yet have representation in the PT item bank.

Table 3.2 shows the total number of items developed per grade level and the high school grade band to accommodate the CAST.

Table 3.2 Total Number of Items Developed per Grade Level and the Grade Band

|  |  |  |  |
| --- | --- | --- | --- |
| Item Type | Grade 5 | Grade 8 | High School |
| Standard discrete item types (non-CR) | 28 | 27 | 27 |
| Discrete CR | 2 | 3 | 3 |
| PT items (4–6 tasks per grade level or grade band) | 72 | 84 | 24 |
| **Totals:** | **102** | **114** | **54** |

The standard discrete item types from table 3.2 included traditional MC items and familiar TEI types (for example, grid, inline choice list).

The operational PTs, which contained four to six items for the CAST, were designed to provide students with an opportunity to demonstrate their ability to apply knowledge and higher-order thinking skills to explore and analyze a complex, real-world scenario.

ETS developed all items for CAST in accordance with the *ETS* *Standards for Quality and Fairness* (2014)across all phases of item and assessment development.

#### Process

Each CAST item is developed through a comprehensive development cycle and designed to conform to principles of quality item writing as defined by the item acceptance criteria that were developed by the CDE and ETS. Further, each item in the CAST item bank is developed to measure a specific PE through integration of the DCI and at least one of the other two dimensions of the CA NGSS—SEPs and CCCs. In addition, guidelines for style and for fairness—including issues related to bias and sensitivity—help item developers and reviewers maintain item consistency across the item bank.

Throughout the item writing process, ETS adheres to its best practices for quality item writing. According to these best practices, item developers conform to the following list of eight attributes for each item:

1. The item is clearly and concisely presented.
2. There is an absence of clueing in the item stem and supporting stimuli.
3. The supporting stimulus, or stimuli, is presented clearly and is construct relevant.
4. There is a single correct answer (for selected-response items only).
5. Distractors are plausible but are incorrect or do not answer the stem (for selected response only).
6. The answer key is correct.
7. The scoring rubric and annotations are accurate, precise, and complete.
8. Item format and content adhere to the principles of universal design.

#### Item Specifications

ETS created item content specifications for CAST using feedback from the CDE and California teachers, with task models guiding the initial development. The item content specifications are extensions of these models intended to be more specific in nature and to incorporate information and feedback gained through the development, review, and administration processes. These specifications describe the characteristics of items that consistently elicit evidence of student mastery of specified aspects of each PE. The specifications were developed in consultation with the CDE, and the CDE determined the emphasis on different aspects of each PE. The specifications include the following:

* Science and Engineering subpractice
* Subpractice assessment targets
* DCI assessment targets
* CCC assessment targets
* Possible phenomena or contexts
* Examples of integration of assessment targets and evidence
* California Environmental Principles and Concepts (where applicable)
* Common misconceptions
* Additional assessment boundaries

In accordance with the iterative nature of ECD described previously, the item content specifications used to produce the CAST items will be updated periodically to support subsequent rounds of item development.

#### Recruitment and Selection of Item Writers

During item development cycles, senior ETS content staff screen applications for CAST item writers, and ETS approves only those with strong content and teaching backgrounds for the item writing training program.

#### Item Writer Training

Item writer training is a vital part of establishing the validity chain for item and task development. In addition to relying on internal item writing experts for CAST, ETS recruited and trained educators in the CA NGSS.

The three primary goals for the training were to

1. provide teachers with knowledge, via professional development on writing items, that they can use to help develop or refine their own classroom teaching and assessments;
2. ensure that teachers who successfully completed the training were ready to develop high-quality items for CAST; and
3. leverage the experiences, perspectives, and expertise of the teachers in writing items for CAST.

ETS held item writer training sessions to provide prospective item writers with professional development in several areas. A review of the general assessment development process gave trainees a sense of the total life cycle of an item.

Participants learned best practices in item writing to provide clarity within the item and avoid bias or sensitivity concerns, learned how to review a passage for item opportunities, and were introduced to how the new, innovative item types work.

Given that the trainees were California educators and educational leaders, ETS also emphasized incorporation of current effective teaching practices and instructional activities. Small-group and individual work generated sample items that the ETS facilitators then used in a large-group discussion to analyze and ascertain overall item quality. The ETS team also provided post hoc feedback via email and phone calls to trained item writers on further item samples and ideas submitted ahead of contractual item submissions.

### ETS Item Review Process

After items were drafted, ETS placed items developed for CAST through an extensive internal item review process designed to provide the best standards-based assessments possible. This section summarizes the item review process that confirmed the quality of CAST items.

#### Overview

Once an item was accepted for authoring, ETS employed a series of internal reviews. These reviews used established criteria to judge the quality of item content and to ensure that each item measured what it was intended to measure. These internal reviews also examined the overall quality of the items ahead of their being reviewed by the CDE and by educators at IRMs, which are described in more detail in section [*3.5 California Educator Review*](#_California_Educator_Review_1).

All items were entered into the Item Banking Information System (IBIS) with corresponding artwork and metadata. Within IBIS, items received content reviews by ETS assessment specialists and fairness and editorial reviews by ETS editors and fairness reviewers.

The CDE reviewed proposed changes to items in response to reviews by the participants of the IRMs to ensure the quality of the item pool. The CDE then gained access to CAST items and conducted reviews in IBIS. ETS revised items in response to comments from the CDE prior to using them in the assessment forms.

The ETS review process for CAST includes the following; these tasks are described in the next subsections:

1. Content review
2. Accessibility review
3. Editorial review
4. Sensitivity and fairness review

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

#### ETS Content Review

On all items ETS developed, content-area assessment specialists conducted three reviews on items and stimuli—accessibility, editorial, and bias and sensitivity. These assessment specialists verified thatthe items and stimuli were in compliance with the ETS written guidelines for clarity, style, accuracy, and appropriateness for California students and were also in compliance with the approved item specifications, the *California Assessment of Student Performance and Progress and English Language Proficiency Assessments for California Item Review Acceptance Criteria* (ETS, 2019a), and other ETS-produced procedures such as the ETS guidelines for fair tests and communications (2016a). Assessment specialists reviewed each item in terms of the following characteristics:

* Relevance to the purpose of the assessment
* Match of each item to the item specifications, including the tier of item complexity
* Match of each item to the principles of quality item writing
* Match of each item to the identified standard or standards
* Difficulty of the item
* Accuracy of the content of the item
* Readability of the item or passage
* Grade-level and grade-band appropriateness of the item
* Appropriateness of any illustrations, graphs, or figures

Assessment specialists verified the classification of each item, both to evaluate the correctness of the classification and to confirm that the task posed by the item was relevant to the outcome it was intended to measure. The reviewers could accept the item and classification as written, suggest revisions, or recommend that the item be discarded. These steps occurred prior to CDE review.

#### ETS Accessibility Review

The ETS Accessible Content & Inclusive Solutions team advised on accessibility of items and item types during the ETS content review. These experts on alternate test formats reviewed all items, with a focus on accessibility for all student populations, and provided potential refinement solutions to improve the accessibility in items and assessments.

#### ETS Editorial Review

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

#### ETS Sensitivity and Fairness Review

ETS assessment specialists who were specially trained to identify and edit or eliminate items that contained content or wording that could be construed to be offensive to, or biased against, members of specific student groups (for example, ethnicity, race, or gender) conducted the next level of review (ETS, 2014, 2016). These trained staff members reviewed every item before the CDE and IRMs. Newly developed items were then submitted to the CDE for review prior to educator reviews.

The review process promoted a general 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
* Topics that may be unsettling or otherwise distract the student from the content being measured, such as natural disasters, disease, or family discord
* Contributions of diverse groups (including ethnic and minority groups, individuals with disabilities, and women) to the history and culture of the United States and the achievements of individuals within these groups
* Item accessibility for language learners of diverse backgrounds

### California Department of Education Review

After ETS reviews of items were completed, the items were reviewed by the CDE content teams. CDE content experts reviewed the items using the same criteria used in the ETS reviews. After CDE reviews occurred, ETS made edits to the items based on the CDE feedback, and the items were then finalized for IRMs with California educators.

### California Educator Review

#### California Educators as Content Experts

In addition to the ETS internal reviews, meetings with California educators were held at the end of the item review process as the final content-expert review that items must undergo before being field-tested on the CAST. The California educators fill an advisory role to the CDE and ETS and provide guidance on matters related to item development for CAST.

These educators were responsible for reviewing all newly developed items for alignment with the CA NGSS. Meeting participants also reviewed the items for accuracy of content, clarity of phrasing, and overall quality. In their examination of items, participants could raise concerns related to the accuracy of the science content and grade-level appropriateness as well as gender, racial, ethnic, or socioeconomic bias.

#### Composition of Item Review Panels

The panelists for the IRMs for CAST items included current and former teachers, resource specialists, administrators, curricular experts, and other education professionals. Minimum qualifications to be invited to participate are

* three or more years of general teaching experience in kindergarten through grade twelve;
* three or more years of teaching experience in science;
* bachelor’s or higher degree in science or education; and
* knowledge of, and experience with, the CA NGSS.

School administrators; LEA, county content, or program specialists; or university educators met the following qualifications to be invited to participate:

* Three or more years of experience as a school administrator, LEA, county content, or program specialist; or university instructor in a grade-specific area or area related to science
* Bachelor’s or higher degree in a grade-specific or content area related to science
* Knowledge of, and experience with, the CA NGSS

Every effort was made to ensure that groups of item reviewers included a wide representation of genders, geographic regions, and ethnic groups in California.

Item reviewers were recruited by the Sacramento Office of Education through an online application process. Recommendations were solicited from LEAs and county offices of education as well as from CDE and SBE staff. ETS assessment directors reviewed applications and confirmed that an applicant’s qualifications met the specified criteria. Applicants who met the criteria had their information forwarded to CDE staff for further review and agreement before invitations to participate were distributed.

#### Meetings for Review of CAST Field Test Items

The 2022–23 CAST IRMs were held virtually in 2023: from January 17 to 25 for grade eight, January 24 to 25 for high school, and March 15 to 21 for grade five. ETS content-area assessment specialists facilitated the IRMs. Each meeting began with a brief training session on how to review and make recommendations for accepting or revising items.

ETS provides training on the following topics:

* Overview of the purpose and scope of CAST
* Overview of the CAST design specifications
* Overview of criteria for evaluating test items
* Review and evaluation of items for fairness concerns

The criteria for reviewing items include the following:

* Overall technical quality
* Alignment with the PEs
* Alignment with the construct being assessed by the standard
* Difficulty range
* Clarity
* Correctness of the answer
* Plausibility of the distractors
* Bias and sensitivity factors

ETS provides guidelines for reviewing items, which the CDE approves. The set of guidelines for reviewing items is summarized as follows:

* Does the item

have one and only one clearly correct answer or set of correct answers?

measure the achievement standard?

align with the construct being measured?

test worthwhile concepts or information?

* Is the stimulus, if any, for the item

required to answer the item?

likely to be interesting to students?

clearly and correctly labeled?

providing all the information needed to answer the item?

Once ETS staff compile and review the panel’s feedback, the feedback is delivered to the CDE for further review and guidance on decisions on whether to field-test the items.

#### Data Review Meeting

After items were included in a field test segment administered to students, ETS prepared items that did not meet all statistical thresholds for review by the CDE and California educators on June 18 and 20, 2024.

Educators selected to participate in the DRM met the same qualifications as those selected for item review panels. Educators who were part of the data review panel were assigned a training video in Upskill—a centralized, online location for training materials—to give them an overview of what is involved in a DRM as well as an understanding of the statistical measures used to review the flagged field test items. This was followed by ETS conducting an introductory training for reviewers to highlight any new issues and serve as a statistical refresher at the beginning of the DRM. Reviewers then made recommendations about which field test items should be included in the item bank for future form assembly. If a field test item was considered problematic and not to be included in the item bank, it would be either removed from the bank or revised and once again follow the steps in the item development process, including field testing. ETS psychometric and content staff were available to reviewers throughout this process.

Content staff facilitated the meeting, confirming that all educators weighed in on each flagged field test item to confirm there were no concerns, from a content perspective, as it pertained to the item flag. ETS content staff and psychometricians provided training on the item statistics and responded to questions about the item statistics during the item discussion. The DRM participants reviewed the content and statistics of each item and then made a recommendation to accept or reject an item.

Content staff recorded each participant’s recommendations and comments regarding the flagged field test items. The feedback was referenced when working with the CDE to reconcile educator feedback and to make a final decision on whether to include each field test item in the operational pool.

Considering all field-tested items (that is, flagged and unflagged), grade five had 98 percent acceptance into the operational item bank (94 of 96 field-tested items), grade eight had 94 percent acceptance (102 of 108 field-tested items), and the high school grade band had 95 percent acceptance (57 of 60 field-tested items).

Table 3.3, which contains the results of the DRM, shows the following information:

* Field test items analyzed by ETS psychometricians
* Items flagged and reviewed at the DRM
* Flagged, field-tested items accepted without edits at the DRM
* Flagged, field-tested items rejected outright at the DRM
* Overall number and percentage of field test items that were accepted into the operational item bank after the DRM concluded

Table 3.3 Item Data Review Results

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Grade Five** | **Grade Eight** | **High School** |
| Total items field-tested | 96 | 108 | 60 |
| Items not flagged for the DRM | 86 | 98 | 45 |
| Items flagged to bring to the DRM | 10 | 10 | 15 |
| Items accepted at the DRM | 8 | 4 | 12 |
| Items rejected at the DRM | 2 | 6 | 3 |
| Field-tested items accepted into the operational bank | 94 | 102 | 57 |
| Percentage of field-tested items accepted into the operational bank | 98% | 94% | 95% |

### References

Achieve. (2015). Next Generation Science Standards evidence statements.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

California Department of Education. (2016). *California’s Next Generation Science Standards (CA NGSS) assessment plan*. Presented to the California State Board of Education in March 2016. California Department of Education website.

California Department of Education. (2019). *California Science Test assessment fact sheet.* California Department of Education website.

Educational Testing Service. (2014). *ETS standards for quality and fairness*. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2016a). *ETS guidelines for fair tests and communications*. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2016b). *Proposed design for California’s Next Generation Science Standards general summative assessments* [Unpublished report]. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2019a). *CAASPP and ELPAC item acceptance criteria* [Unpublished manuscript]. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2019b). *Use of evidence-centered design in CAST item and test development* [Unpublished report]. Princeton, NJ: Educational Testing Service.

Instructional Management Systems (IMS). (2020). *IMS question & test interoperability specification*.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* (ETS Research Report RR-03-16). Princeton, NJ: Educational Testing Service.

## Test Assembly

This chapter discusses the detailed procedures of test assembly for the 2023–24 California Science Test (CAST) administration.

### Overview

ETS assessment specialists and the psychometric team assembled CAST, which was reviewed and approved by the California Department of Education (CDE). This process began with the review and approval of the form assembly specifications by the CDE. These specifications described the content characteristics, psychometric characteristics, and quantity of items to be used in the operational 2023–24 CAST.

#### Design

CAST design is based on the California State Board of Education (SBE)–approved high-level test design for an operational assessment, which requires that students in the tested grade levels participate in three segments of the assessment: Segment A, Segment B, and Segment C. Segments A and B contribute to individual student score reporting. Segment C is used for field testing, where the items do not count toward students’ score reporting. The braille computer-based forms and paper–pencil tests (PPTs) did not include field test items because of the increase in testing time needed by the students using the braille resource and because the same PPT forms are used over multiple test administrations. Additionally, very few students test using either braille or paper forms and so are not necessarily representative of the overall testing population.

In the 2023–24 test administration, Segment A contained two blocks of discrete items equally distributed between the two blocks. All students being assessed at a specific grade level or the high school grade band (that is, grade ten, eleven, or twelve) received the same two Segment A blocks. The delivery order of the two Segment A blocks was random, with either one being presented first to the student.

Segment B included three performance tasks (PTs) from the pool. Each PT consisted of five to six items worth six to seven points total. Through analysis of the items, each PT was identified as applying to one of the three science domains. Where a PT had items that were aligned with an Engineering, Technology, and Applications of Science (“ETS” in table 4.A.1 through table 4.A.3 in [appendix 4.A](#_Appendix_4.A:_Test_1)) performance expectation (PE), a science domain for the item was designated on the basis of the context of the item. Where a PT had items that were part of more than one science domain, a primary content domain was designated on the basis of the predominant domain in the storyline. Each student received one PT from each of three different domains—Earth and Space Sciences, Life Sciences, and Physical Sciences (“ESS,” “LS,” and “PS” in table 4.A.1 through table 4.A.6), for a total of three PTs per student. The PTs were presented after the discrete blocks and were delivered randomly.

Segment C included a block of discrete field test items or a field test PT from the pool. The field test pool contained four or six discrete item blocks and various PT blocks at each grade level or the high school grade band: 12 for grade five, 14 for grade eight, and 4 for high school. Each student was administered one block with six discrete items or one PT with six items. All of the 30 field test PT blocks came from parallel PTs. Parallel PTs were field-tested as two distinct Segment C PTs with identical stimuli but unique item sets. One goal of parallel PTs was to increase the likelihood of yielding an operational Segment B PT from among the pool of parallel PT field test items. The operational Segment B PT would align with the SBE-approved CAST blueprint and CDE guidelines.

Table 4.1 lists the total number of unique items available for administration per segment across the forms for each grade level or the grade band. (Braille computer-based forms and PPTs did not include Segment C and had three PTs instead of six PTs available.) While these totals reflect the items available in the 2023–24 summative general operational forms, the number of items with which a single student interacts is described in subsection [*2.1.1 Design Guidelines*](#_Design_Guidelines).

Table 4.1 Number of Unique Items Assessed on the CAST

|  |  |  |  |
| --- | --- | --- | --- |
| **Segment** | **Grade 5** | **Grade 8** | **High School** |
| A (2 discrete blocks) | 26 | 28 | 32 |
| B (6 PT blocks) | 35 | 33 | 35 |
| C (4–6 discrete blocks) | 24 | 24 | 36 |
| C (4–14 PT blocks) | 72 | 84 | 24 |

### Automated Test Assembly

Automated test assembly (ATA) methods were used to assemble the first version of Segment A of the operational form for each CAST grade level. ATA is an algorithmic process by which the optimal set of items can be selected subject to a set of constraints. An ATA user defines optimal test assembly parameters and specifies the appropriate constraints.

For test assembly purposes, “optimal” is usually defined as providing the highest level of test information at one or more predetermined, theoretically important values of the latent trait variable, theta. For CAST, a theta value of 0 was used, corresponding to the approximate center of the latent trait distribution. In other words, measurement precision was optimized for test takers of average ability as measured by the assessment. The constraints are defined by blueprint requirements; psychometric considerations (for example, average form-level difficulty); and form refresh requirements.

ATA ensures that ETS is providing the best possible test forms in terms of psychometric performance while still meeting all blueprint requirements.

### Test Blueprint and Other Content Specifications

#### Test Blueprint

Table 4.A.1 through table 4.A.6 in [appendix 4.A](#_Appendix_4.A:_Test_1) show the CAST blueprint approved by the California SBE in January 2020 and used to build the 2023–24 operational forms. For details on the test blueprint for CAST, refer to the *Revised California Science Test Blueprint* (CDE, 2020a).

In these tables, an asterisk (\*) indicates that, across the three science content domains, a student will receive zero to one item assessing Engineering, Technology, and the Applications of Science. The item(s) may be discrete or part of a PT. Two asterisks (\*\*) indicate that the CAST item specifications provide greater detail on the assessment targets by PE (CDE, 2020b).

The test blueprint also specifies the PE distribution for Segment A items by the disciplinary core ideas (DCIs) (within each content domain), science and engineering practices (SEPs), and crosscutting concepts (CCCs). These tables are included in [appendix 4.B](#_Appendix_4.B:_Performance_1) as figure 4.B.1 through figure 4.B.3.

Segment A is designed to assess a student’s mastery of a breadth of PEs of the California Next Generation Science Standards (CA NGSS) in the grade levels and high school grade band tested.

The tables display an “X” for the intersections of SEPs, DCIs, and CCCs articulated in the PEs. These intersections represent opportunities to develop items that can be used to assemble Segment A. While each individual item reflects the intersection of a SEP, DCI, and CCC, the tables also indicate the proposed distribution of Segment A items by DCI, SEP, and CCC. Segment A on the 2023–24 CAST general online forms had 8 to 11 items in each of the three science domains: Earth and Space Sciences, Life Sciences, and Physical Sciences. Zero to three items assessed the Engineering, Technology, and Applications of Science PEs, but for scoring and reporting purposes, items written to those PEs were assigned to one of the three science domains depending on the context of their stimulus.

### Test Production Process

The final steps in production of CAST are to identify, select, and review items. These are discussed in the following subsections.

#### Content Criteria

Operational segments A and B in 2023–24 were built to be parallel in terms of other content and statistical specifications. Operational segments A and B in 2023–24 were refreshed from the 2022–23 test administration to meet refresh requirements and other specifications that are based on the blueprint.

Segment A had an item type distribution of 43 to 50 percent multiple-choice items and 42 to 46 percent technology-enhanced items per form. Each Segment A block had one to three constructed-response (CR) items. Each Segment B PT contained one CR item.

Cognitive complexity as measured by depth of knowledge (DOK) had a distribution for Segment A where up to 7 percent of the items on the form had DOK 1, 43 to 50 percent measured at DOK 2, and 44 to 54 percent measured at DOK 3 or 4. DOK values were determined and reviewed by educators, ETS assessment specialists, and CDE staff periodically during item development and item review meetings (IRMs).

For more information on how the additional field test items were selected to be added, refer to subsection [*3.5.3 Meetings for Review of CAST Field Test Item*s](#_Meetings_for_Review).

For a typical operational test production process when test forms are not being reused, assessment developers identified discrete items to be removed from the previous test administration’s operational Segment A, and then ATA was used to select the best replacement discrete items from the eligible discrete item pool. For operational Segment B, assessment developers select the best replacement PTs from the eligible PT item pool. For both segments, assessment developers and ATA procedures ensure that test forms, as a whole,

* meet the coverage specifications of the test blueprint (subsection [*4.3.1 Test Blueprint*](#_Test_Blueprint)),
* meet item selection criteria developed by the ETS psychometrics team (subsection [*4.4.3 Psychometric Review*](#_Psychometric_Criteria_and)),
* represent a wide variety of item types, and
* provide a wide variety of item contexts.

Refer to section [*4.2 Automated Test Assembly*](#_Automated_Test_Assembly_1) for additional discussion of ATA.

#### Test Forms

Table 4.2 provides the number of blocks for each grade level and the high school grade band for each segment of the general computer-based form.

Table 4.2 The Number of Blocks for Each Segment for 2023–24 CAST General Computer-‍based Forms

|  |  |  |  |
| --- | --- | --- | --- |
| **Block Description** | **Grade 5** | **Grade 8** | **High School** |
| Operational Discrete Blocks | 2 | 2 | 2 |
| Operational PT Blocks | 6 | 6 | 6 |
| Field Test Discrete Blocks | 4 | 4 | 6 |
| Field Test PT Blocks | 12 | 14 | 4 |

The number of blocks summarized in table 4.2 resulted in the following number of unique forms for the general computer-based administration:

* Eight forms for grade five
* Eight forms for grade eight
* Eight forms for high school (grade ten, eleven, or twelve)

By including field test blocks, the total numbers of unique forms are the following:

* One hundred twenty-eight forms for grade five
* One hundred forty-four forms for grade eight
* Eighty forms for high school (grade ten, eleven, or twelve)

In addition to the forms for the general computer-based administration, there was also one computer-based braille form and one paper–pencil form available in the 2023–24 test administration. Each of these two forms included two discrete blocks in Segment A and three PTs in Segment B. No field test blocks were included for either of these two forms.

All the forms mentioned previously were evaluated using the psychometric criteria as described in subsection [*4.4.3 Psychometric Review*](#_Psychometric_Criteria_and). The variation in the number of forms across grade levels and the grade band is partially due to the number of operational and field test discrete and PT item blocks available for each grade level or the grade band.

#### Psychometric Review

Classical item statistics such as the *p*-value (item difficulty), item-total polyserial correlation (item discrimination), and the item response theory (IRT) parameters obtained from previous CAST administrations were used to inform the item selection for the operational forms used in the 2023–24 test administration. Refer to subsection [*8.2.1 Classical Item Difficulty Indices (p-value)*](#_Classical_Item_Difficulty), subsection [*8.2.2 Item-Total Correlation*](#_Item-Total_Correlation_1), and section [*8.4 Item Response Theory Analyses*](#_Item_Response_Theory) for more details on these criteria.

The item selection guidelines listed next suggest the most favorable qualities for items used in an operational form. There may be items in the pool that do not meet these criteria but have been fully reviewed to ensure there are no content issues. Such items may be used in the operational form only when they are needed for content coverage.

The following guidelines were used to select items:

* All items were operationally ready, with item statistics.
* All items conformed to the specifications in the test blueprint.
* Items with *p*-values between 0.20 and 0.95 were used. A *p*-value less than 0.20 suggested that the item might be too difficult; a *p*-value greater than 0.95 for a dichotomous item or greater than 0.80 for a polytomous item suggested that the item might be too easy.
* Items that were too easy or too difficult were not used, as they provided little information on evaluating students’ abilities.
* The item-total polyserial correlation was at least 0.20.
* Items flagged for C-DIF were not used unless it was necessary for content coverage (refer to subsection [*8.3.3 Classification*](#_Classification_1) for more details on the differential item functioning [DIF] classification). All items flagged for C-DIF were reviewed by a DIF panel that included members of the focal groups that were affected and who confirmed the items were not biased before the items could be selected for use. The panelists did not have a vested interest in the outcome of the decision.
* Item discrimination IRT *a-*parameter was positive with a standard error (SE) at 0.3 or less.
* Item difficulty IRT *b*-parameter was in the range of −4 and 4, with an SE at 0.3 or less.
* Polytomous items should have item category IRT *d-*parameter in the range of −4 and 4, with SE at 0.3 or less. The distance between two adjacent step parameters should be 2.5 or less.

At the form level, the blocks in Segment A should also meet the following statistical specifications:

* The average *b*-value of the form should be around zero. The desired item difficulty distribution for the two Segment A blocks is shown in table 4.3. The item difficulty distribution should be even across domains. However, because of the size of the item pool and the content constraints, it might be difficult to meet this distribution exactly, especially at the domain level. When that happens, the content constraint should be given a higher priority. (Refer to table 8.E.4 through table 8.E.6 in [appendix 8.E](#_Appendix_8.E:_Item) for item difficulty distribution by content domain.)

Table 4.3 Desired Item Difficulty Distribution in Segment A

|  |  |  |
| --- | --- | --- |
| ***b-p*arameter Range** | **Percent of Items** | **Number of Items** |
| b ≤ −1.5 | 7% | 2–3 |
| −1.5 < b ≤ −0.5 | 24% | 7–8 |
| −0.5 < b ≤ 0.5 | 38% | 12–13 |
| 0.5 < b ≤ 1.5 | 24% | 7–8 |
| b > 1.5 | 7% | 2–3 |

* The two blocks in Segment A should have parallel test information functions. For each block, test information should be maximized around three theta points where the cut scores to reach each achievement level were set. Standard setting following the 2018–19 test administration resulted in the theta cut scores shown in table 4.4.

Table 4.4 Theta Cut Scores Based on Standard Setting Results

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Level 2 Theta** | **Level 3 Theta** | **Level 4 Theta** |
| Grade 5 | −0.96 | 0.51 | 1.35 |
| Grade 8 | −1.01 | 0.55 | 1.47 |
| High school | −1.11 | 0.55 | 1.68 |

* The test characteristic curve from the forms built should be comparable to those from the past test administrations, especially at the ability intervals with high density.
* The nature of CAST test design poses considerable challenges to ensuring that the sequence of items is similar between field testing and operational administration. However, to the extent possible, effort should be made to sequence items on the operational form as closely as possible to the sequence in the blocks when they were field-tested or their positions in the blocks when they were tested during a prior operational administration.

ETS assessment developers sent the proposed forms—both the general forms and the braille forms—to the ETS psychometrics team for approval. For the operational forms, the psychometric review included a check on whether the selected items and the forms met the test blueprint as described in table 4.A.1 through table 4.A.6 in [appendix 4.A](#_Appendix_4.A:_Test_1) and the psychometric criteria as described in this subsection.

The psychometric review is an iterative process. The psychometrics team worked with the content team to find replacement items, if items or forms were identified as not meeting the specifications; forms review continued until all forms met the specifications. The number of blocks reviewed is summarized in table 4.5. Note that the number of operational discrete blocks and PTs includes the number of blocks included in the computer-based braille form and the paper–pencil form as well as forms used in the general computer-based assessment.

Table 4.5 Number of Blocks Reviewed Across All Forms

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Number of Operational Discrete Blocks** | **Number of Operational PT Blocks** | **Number of Field Test Discrete Blocks** | **Number of Field Test PT Blocks** | **Total Number of Blocks** |
| Grade 5 | 6 | 12 | 4 | 12 | 34 |
| Grade 8 | 6 | 12 | 4 | 14 | 36 |
| High school | 6 | 12 | 6 | 4 | 28 |

#### Content Review of Forms

After psychometric approval, the proposed assessment underwent two additional content reviews and one editorial review. The content reviewers were assessment developers who had not worked on developing the test forms they were reviewing. These reviewers brought a fresh perspective to the review. They were given the appropriate materials and documentation to complete the following tasks:

* Verification of item keys
* Identification of possible clueing across the items
* Verification that individual items aligned with the CA NGSS
* Verification of coverage of the CA NGSS
* Identification of any possible grammatical or production errors

#### California Department of Education Forms Review

The CDE used a gatekeeper process to review all test materials. Test materials for review and approval by the CDE included form planners and student-facing items in the Test Delivery System (TDS). All test materials were approved before they were made available for use.

For the reviews of form planners, ETS initiated the review by submitting materials to the CDE via the gatekeeper system, along with the criteria for the review. CDE consultants performed the initial review and returned comments and requests for revisions to ETS. ETS staff then revised the materials as requested and returned them to the CDE consultants, who reviewed the updated materials. If the test materials needed additional revisions, they were returned to ETS for further modifications.

Once CDE consultants found that the test materials met the review criteria, the CDE consultants submitted the test materials to the CDE administrator for approval. Test materials that were approved with revisions were revised by ETS and resubmitted for approval. Test materials that were not approved needed significant revisions and had to be submitted to the consultants again before they could be resubmitted to the CDE administrator for approval.

Following the ETS content review, all proposed assessments were sent to the CDE for review via the Item Banking Information System Content Review Tool to ensure the proposed assessments met the CAST blueprint requirements and to check for possible statistical issues or clueing between items. The CDE was provided with block builders that catalogued information and documented resolutions for staff comments about the assembled segments.

Comments from the CDE to make changes to the forms were acted upon by the ETS assessment development team.

#### Configuration of the Test Delivery System

Once all the test reviews were completed and concerns, if any, had been resolved, the official ordered item sequence of the proposed forms was sent to Cambium Assessment, Inc. (CAI) for configuration of the TDS. Unlike other stages of the test production process, this stage must occur prior to every administration of CAST, even in the case of a form reuse.

Each item underwent an extensive platform review on different operating systems, such as Windows, Linux, and iOS, to ensure that the item’s appearance was consistent across all platforms.

The platform review was conducted by a team at CAI consisting of a team leader and several team members. The team leader presented the item as it was approved in ETS and CAI item banks. Each team member was assigned a different platform—hardware device and operating system—and reviewed the item to see that it rendered as expected. This platform review meeting ensured that all items were presented consistently to all students regardless of testing device or operating system for standardization of the test administration.

Prior to operational deployment, the testing system and content were deployed to a staging server where they were subject to user acceptance testing (UAT) by both ETS and CAI staff. The TDS UAT served as both a software evaluation and a content approval.

Following the UAT by ETS and CAI staff, separate UAT cycles were conducted by the CDE. The UAT review provided the CDE with an opportunity to interact with the exact assessment that would be administered to the students. The CDE had to approve the CAST UAT before the assessment could be released for administration to students.

#### Test Form Delivery

Students were administered the two operational, discrete blocks first, in Segment A, followed by the three operational PT blocks in Segment B. Within each segment (that is, Segment A and Segment B), the blocks were delivered in a random order. Segment C, which is used for field testing, is a block that includes six discrete items or a PT.

### Performance Expectation Coverage

The various blocks of items that compose each segment of CAST covered an extensive range of PEs; these PEs at the operational item pool–level are shown for the grade levels and the high school grade band in table 4.6.

Table 4.6 PEs Assessed on the 2023–24 CAST

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Assessed PEs** | **Grade-Level or Grade-Band PEs** | **Percent of Assessed PEs** |
| Grade 5 | 37 | 45 | 82 |
| Grade 8 | 39 | 59 | 66 |
| High school | 40 | 71 | 56 |

Test forms were built to rotate assessed PEs in segments A and B as much as possible so that all PEs can be assessed operationally over the course of a three-year period, as noted in the test blueprint. The 2023–24 CAST forms represent the second year of the three-year period; as such, table 4.7, which shows cumulative PE coverage for the three-year period, will be completed after two further CAST administrations. (In the table, “N/A” indicates where percentages will be provided after future test administrations. The sum of percentages from three operational administrations will be greater than 100 percent because some of the same PEs will be assessed in multiple test administrations.)

Table 4.7 Cumulative Coverage of PEs Over Three Years

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Percent of PEs Assessed in Year One** | **Percent of PEs Assessed in Year Two** | **Percent of PEs Assessed in Year Three** | **Percent of Unique PEs Assessed in Years One and Two** |
| Grade 5 | 76 | 82 | N/A | 89 |
| Grade 8 | 64 | 66 | N/A | 81 |
| High school | 54 | 56 | N/A | 70 |

### Special Version Forms

#### Braille Form

ETS assembled braille forms for students with visual impairments. The same operational pool of items and PTs that were used for embedded designated supports and accommodations on the form for the general population were also used for the grade five, grade eight, and high school braille online forms.

Given the need to evenly distribute science content, item types, and such within Segment A blocks, only some Segment A items appeared in the same or similar positions on the braille form as they did in Segment A of the form for the general population. The braille form Segment B included three PTs, one from each of the three different science content domains. These three PTs were selected from PTs available for Segment B of the general population form.

If an item that relied heavily on visual input—whether through item type or visual stimuli—was needed to meet the blueprint, the item was either adapted or “twinned” to meet the accessibility needs of the population of students with visual impairments. Whether items were adapted or twinned, the item construct and overall cognitive complexity was maintained as closely as possible with the original parent item.

Adaptation, which did not change the item type, may have included simplified graphics, more descriptive alternative text for images, or other changes to make the item more accessible to refreshable braille devices, embossed tactile graphics, or screen readers. Twinning an item meant the item used a different item type while maintaining the same construct and storyline of the original item.

Techniques and methods for adapting or twinning items were developed through a cognitive laboratory study with students with visual impairments, IRMs with teachers of the visually impaired (TVIs), and a comparability study with blind or visually impaired scientists and TVIs.

#### Paper–Pencil Form

The 2023–24 CAST administration included a PPT form for students whose individualized education program (IEP) or Section 504 plan specified testing on a paper–pencil form, or when a school experienced unexpected, temporary technology issues beyond the school’s control.

The grade five and high school PPT forms used in 2023–24 were the same as the 2022–23 PPT form and complied with item selection and forms construction criteria noted in prior sections. The grade eight PPT form used in 2023–24 was refreshed from the 2022–23 PPT form. Three discrete items that complied with item selection and forms construction criteria noted in prior sections were refreshed on the grade eight PPT form. Standard, large-print, and braille paper–pencil forms were developed.

#### Forms with Accessibility Features Other Than Braille

All of the general form blocks were used to provide accessible content for those students who had been assigned one or more designated supports or accommodations, as determined by an educator, IEP team, or Section 504 plan. Items were embedded with content for text-to-speech, stacked Spanish, translation glossaries, and American Sign Language videos. Refer to [*5.4.1 Accessibility Resource Categories*](#_Accessibility_Resource_Categories_1) for a list of designated supports and accommodations available during the 2023–24 CAST administration.

Both Segment A blocks and all Segment B PTs found on the general forms included designated supports and accommodations; a student received one Segment B PT for each science domain. All discrete item blocks and PT blocks from different science domains in Segment C were designated as accessible with these resources. Students received either the discrete item block or one PT for Segment C.

### References

California Department of Education. (2020a). *CAST item specifications*. California Department of Education website.

California Department of Education. (2020b). *Revised California Science Test blueprint*.California Department of Education website.

### Appendix 4.A: Test Blueprint

In table 4.A.1 through table 4.A.6, an asterisk (\*) indicates that, across the three science content domains, a student will receive zero to one items assessing Engineering, Technology, and the Applications of Science (ETS). The item(s) may be discrete or part of a PT. Two asterisks (\*\*) indicate that the CAST Item Specifications provide greater detail on the assessment targets by PE (CDE, 2020).

Table 4.A. CAST Blueprint for Segments Contributing to Individual Scores from the Earth and Space Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Science Content Domain and DCI\*\*** | **Segment A: Discrete Items by DCI—Grade 5** | **Segment A: Discrete Items by DCI—Grade 8** | **Segment A: Discrete Items by DCI—High School** | **Segment B: PTs** |
| **ESS1:** Earth’s Place in the Universe | 1–2 | 1–3 | 1–5 | 0–1 PTs for all ESS DCI strands |
| **ESS2:** Earth’s Systems | 1–5 | 1–5 | 1–6 | 0–1 PTs for all ESS DCI strands |
| **ESS3:** Earth and Human Activity | 1–3 | 1–4 | 1–5 | 0–1 PTs for all ESS DCI strands |
| **ETS1:** Engineering Design | \* | \* | \* | 0–1 PTs for all ETS (ESS) DCI strands |
| **Total for ESS:** | **8–9 items** | **8–10 items** | **9–12 items** | **4–6 items per PT** |

Table 4.A. CAST Blueprint for Segments Contributing to Individual Scores from the Life Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Science Content Domain and DCI\*\*** | **Items by DCI in Segment A—Grade 5** | **Items by DCI in Segment A—Grade 8** | **Items by DCI in Segment A—High School** | **Segment B: PTs** |
| **LS1:** From Molecules to Organisms: Structures and Processes | 1–2 | 1–6 | 1–6 | 0–1 PTs for all LS DCI strands |
| **LS2:** Ecosystems: Interactions, Energy, and Dynamics | 1–2 | 1–4 | 1–7 | 0–1 PTs for all LS DCI strands |
| **LS3:** Heredity: Inheritance and Variation of Traits | 1–2 | 1–2 | 1–2 | 0–1 PTs for all LS DCI strands |
| **LS4:** Biological Evolution: Unity and Diversity | 1–4 | 1–5 | 1–5 | 0–1 PTs for all LS DCI strands |
| **ETS1:** Engineering Design | \* | \* | \* | 0–1 PTs for all ETS (LS) DCI strands |
| **Total for LS:** | **8–9 items** | **8–10 items** | **9–12 items** | **4–6 items per PT** |

Table 4.A. CAST Blueprint for Segments Contributing to Individual Scores from the Physical Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Science Content Domain and Disciplinary Core Idea (DCI)\*\*** | **Items by DCI in Segment A—Grade 5** | **Items by DCI in Segment A—Grade 8** | **Items by DCI in Segment A—High School** | **Segment B: PTs** |
| **PS1:** Matter and Its Interactions | 1–3 | 1–5 | 2–7 | 0–1 PTs for all PS DCI strands |
| **PS2:** Motion and Stability: Forces and Interactions | 1–4 | 1–4 | 1–5 | 0–1 PTs for all PS DCI strands |
| **PS3:** Energy | 1–4 | 1–4 | 1–4 | 0–1 PTs for all PS DCI strands |
| **PS4:** Waves and Their Applications in Technologies for Information Transfer | 1–2 | 1–2 | 1–4 | 0–1 PTs for all PS DCI strands |
| **ETS1:** Engineering Design | \* | \* | \* | 0–1 PTs for all ETS (PS) DCI strands |
| **Total for PS:** | **8–9 items** | **8–10 items** | **9–12 items** | **4–6 items per PT** |

Table 4.A. CAST Blueprint for Segments Contributing to Individual Scores for Grade Five

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Science Content Domain\*\*** | **Segment A: Discrete Items by Science Content Domain** | **Segment A: Number of Points by Science Content Domain** | **Segment B: Number of PTs by Science Content Domain** | **Segment B: Number of Items per PT by Science Content Domain** | **Segment B: Number of Points by Science Content Domain** |
| **ESS** | 8–9 items | 8–12 points | 1 PT | 4–6 items | 6–7 points |
| **LS** | 8–9 items | 8–12 points | 1 PT | 4–6 items | 6–7 points |
| **PS** | 8–9 items | 8–12 points | 1 PT | 4–6 items | 6–7 points |
| **Total:** | **26 items** | **28–32 points** | **3 PTs** | **12–18 items** | **18–21 points** |

Table 4.A. CAST Blueprint for Segments Contributing to Individual Scores for Grade Eight

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Science Content Domain\*\*** | **Segment A: Discrete Items by Science Content Domain** | **Segment A: Number of Points by Science Content Domain** | **Segment B: Number of PTs by Science Content Domain** | **Segment B: Number of Items per PT by Science Content Domain** | **Segment B: Number of Points by Science Content Domain** |
| **ESS** | 8–10 items | 8–13 points | 1 PT | 4–6 items | 6–7 points |
| **LS** | 8–10 items | 8–13 points | 1 PT | 4–6 items | 6–7 points |
| **PS** | 8–10 items | 8–13 points | 1 PT | 4–6 items | 6–7 points |
| **Total:** | **28 items** | **30–34 points** | **3 PTs** | **12–18 items** | **18–21 points** |

Table 4.A. CAST Blueprint for Segments Contributing to Individual Scores for High School

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Science Content Domain\*\*** | **Segment A: Discrete Items by Science Content Domain** | **Segment A: Number of Points by Science Content Domain** | **Segment B: Number of PTs by Science Content Domain** | **Segment B: Number of Items per PT by Science Content Domain** | **Segment B: Number of Points by Science Content Domain** |
| **ESS** | 9–12 items | 9–15 points | 1 PT | 4–6 items | 6–7 points |
| **LS** | 9–12 items | 9–15 points | 1 PT | 4–6 items | 6–7 points |
| **PS** | 9–12 items | 9–15 points | 1 PT | 4–6 items | 6–7 points |
| **Total:** | **32 items** | **34–38 points** | **3 PTs** | **12–18 items** | **18–21 points** |

### Appendix 4.B: Performance Expectation Distribution for Segment A

#### Blueprint

For scoring and reporting purposes, items written to assess PEs associated with Engineering, Technology, and Application of Science (ETS) will be assigned to one of the three science content domains, depending upon the context of their stimulus.

##### Notes on figure 4.B.1:

* X indicates that there is at least one PE at the given intersection of the three dimensions that can be sampled on a test form for Segment A.
* N/A indicates there is no CCC for at least some of the PEs in the column.
* SEPs 1 and 6 have separate components for science and engineering (SEP 1E and SEP 6E). All other SEPs incorporate the same components for both science and engineering.

The California Next Generation Science Standards (CA NGSS) call out the distinctive purposes of practices primarily in two specific SEPs: SEP 1 and SEP 6. For SEP 1 in science (SEP 1), the practice focuses on identifying questions about phenomena. For SEP 1 in engineering (SEP 1E), the practice focuses on defining a problem to be solved. For SEP 6 in science (SEP 6), the goal of the practice is to construct logically coherent explanations of phenomena to incorporate students’ current understanding of science. For SEP 6 in engineering (SEP 6E), the goal is to propose design solutions to balance competing criteria of desired functions.

Refer to the [*Alternative Text for Figure 4.B.1*](#_Alternative_Text_for_55) for a description of this spreadsheet image.

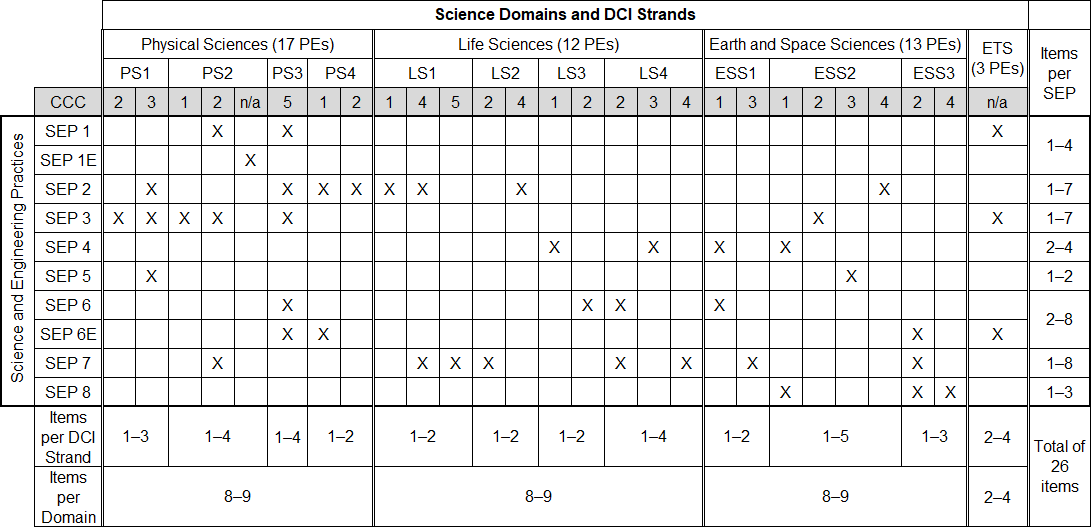


Figure 4.B. PE distribution for Segment A of the CAST grade five assessment

##### Notes on figure 4.B.2:

* X indicates that there is at least one PE at the given intersection of the three dimensions that can be sampled on a test form for Segment A.
* N/A indicates there is no CCC for at least some of the PEs in the column.
* SEPs 1 and 6 have separate components for science and engineering (SEP 1E and SEP 6E). All other SEPs incorporate the same components for both science and engineering.

The CA NGSS call out the distinctive purposes of practices primarily in two specific SEPs: SEP 1 and SEP 6. For SEP 1 in science (SEP 1), the practice focuses on identifying questions about phenomena. For SEP 1 in engineering (SEP 1E), the practice focuses on defining a problem to be solved. For SEP 6 in science (SEP 6), the goal of the practice is to construct logically coherent explanations of phenomena to incorporate students’ current understanding of science. For SEP 6 in engineering (SEP 6E), the goal is to propose design solutions to balance competing criteria of desired functions.

Refer to the [*Alternative Text for Figure 4.B.2*](#_Alternative_Text_for_56) for a description of this spreadsheet image.



Figure 4.B. PE distribution for Segment A of the CAST grade eight assessment

##### Notes on figure 4.B.3:

* X indicates that there is at least one PE at the given intersection of the three dimensions that can be sampled on a test form for Segment A.
* N/A indicates there is no CCC for at least some of the PEs in the column.
* SEPs 1 and 6 have separate components for science and engineering (SEP 1E and SEP 6E). All other SEPs incorporate the same components for both science and engineering.

The CA NGSS call out the distinctive purposes of practices primarily in two specific SEPs: SEP 1 and SEP 6. For SEP 1 in science (SEP 1), the practice focuses on identifying questions about phenomena. For SEP 1 in engineering (SEP 1E), the practice focuses on defining a problem to be solved. For SEP 6 in science (SEP 6), the goal of the practice is to construct logically coherent explanations of phenomena to incorporate students’ current understanding of science. For SEP 6 in engineering (SEP 6E), the goal is to propose design solutions to balance competing criteria of desired functions.

Refer to the [*Alternative Text for Figure 4.B.3*](#_Alternative_Text_for_59) for a description of this spreadsheet image.

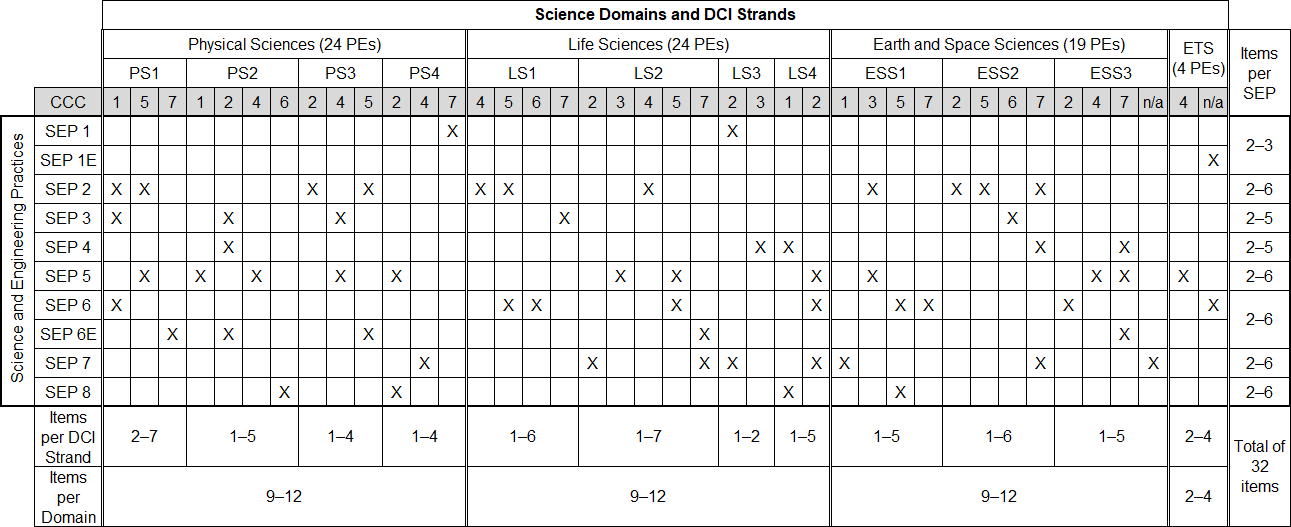


Figure 4.B. PE distribution for Segment A of the CAST high school assessment

#### Accessibility Information

##### Alternative Text for Figure 4.B.1

In the table, an X indicates that there is at least one PE at the given intersection of the three dimensions that can be sampled on a test form for Segment A. The table has an X only in the locations described in the bulleted text that follows for each science domain and the ETS subdomain.

In the Physical Sciences (PS) domain for grade five, there are 17 PEs, organized into four DCI strands, that are distributed among six of the eight SEPs and four of the seven CCCs.

* For the science component of SEP 1, there are at least two PEs.

There is at least one PE in DCI strand PS2, with CCC 2.

There is at least one PE in DCI strand PS3, with CCC 5.

* For the engineering component of SEP 1 (SEP 1E), there is at least one PE in DCI strand PS2, with no CCC.
* For SEP 2, there are at least four PEs.

There is at least one PE in DCI strand PS1, with CCC 3.

There is at least one PE in DCI strand PS3, with CCC 5.

There are at least two PEs in DCI strand PS4, with CCC 1 and CCC 2.

* For SEP 3, there are at least five PEs.

There are at least two PEs in DCI strand PS1, with CCC 2 and CCC 3.

There are at least two PEs in DCI strand PS2, with CCC 1 and CCC 2.

There is at least one PE in DCI strand PS3, with CCC 5.

* For SEP 5, there is at least one PE in DCI strand PS1, with CCC 3.
* For the science component of SEP 6, there is at least one PE in DCI strand PS3, with CCC 5.
* For the engineering component of SEP 6 (SEP 6E), there are at least two PEs.

There is at least one PE in DCI strand PS3, with CCC 5.

There is at least one PE in DCI strand PS4, with CCC 1.

* For SEP 7, there is at least one PE in DCI strand PS2, with CCC 2.

**Alternative Text for Figure 4.B.1 *(continuation one)***

The range of items per DCI strand is described as follows:

* Between one and three items aligned to PEs from DCI strand PS1 will be assessed on Segment A of the CAST.
* Between one and four items aligned to PEs from DCI strand PS2 will be assessed on Segment A of the CAST.
* Between one and four items aligned to PEs from DCI strand PS3 will be assessed on Segment A of the CAST.
* Between one and two items aligned to PEs from DCI strand PS4 will be assessed on Segment A of the CAST.

For the entire PS domain, between eight and nine items will be assessed on Segment A of the CAST.

In the Life Sciences (LS) domain for grade five, there are 12 PEs, organized into four DCI strands, that are distributed among four of the eight SEPs and five of the seven CCCs.

* For SEP 2, there are at least three PEs.

There are at least two PEs in DCI strand LS1, with CCC 1 and CCC 4.

There is at least one PE in DCI strand LS2, with CCC 4.

* For SEP 4, there are at least two PEs.

There is at least one PE in DCI strand LS3, with CCC 1.

There is at least one PE in DCI strand LS4, with CCC 3.

* For the science component of SEP 6 (SEP 6), there are at least two PEs.

There is at least one PE in DCI strand LS3, with CCC 2.

There is at least one PE in DCI strand LS4, with CCC 2.

* For SEP 7, there are at least five PEs.

There are at least two PEs in DCI strand LS1, with CCC 4 and CCC 5.

There is at least one PE in DCI strand LS2, with CCC 2.

There are at least two PEs in DCI strand LS4, with CCC 2 and CCC 4.

The range of items per DCI strand is described as follows:

* Between one and two items aligned to PEs from DCI strand LS1 will be assessed on Segment A of the CAST.
* Between one and two items aligned to PEs from DCI strand LS2 will be assessed on Segment A of the CAST.
* Between one and two items aligned to PEs from DCI strand LS3 will be assessed on Segment A of the CAST.
* Between one and four items aligned to PEs from DCI strand LS4 will be assessed on Segment A of the CAST.

**Alternative Text for Figure 4.B.1 *(continuation two)***

For the entire LS domain, between eight and nine items will be assessed on Segment A of the CAST.

In the Earth and Space Sciences (ESS) domain for grade five, there are 13 PEs, organized into three DCI strands, that are distributed among seven of the eight SEPs and four of the seven CCCs.

* For SEP 2, there is at least one PE in DCI strand ESS2, with CCC 4.
* For SEP 3, there is at least one PE in DCI strand ESS2, with CCC 2.
* For SEP 4, there are at least two PEs.

There is at least one PE in DCI strand ESS1, with CCC 1.

There is at least one PE in DCI strand ESS2, with CCC 1.

* For SEP 5, there is at least one PE in DCI strand ESS2, with CCC 3.
* For the science component of SEP 6, there is at least one PE in DCI strand ESS1, with CCC 1.
* For the engineering component of SEP 6 (SEP 6E), there is at least one PE in DCI strand ESS3, with CCC 2.
* For SEP 7, there are at least two PEs.

There is at least one PE in DCI strand ESS1, with CCC 3.

There is at least one PE in DCI strand ESS3, with CCC 2.

* For SEP 8, there are at least three PEs.

There is at least one PE in DCI strand ESS2, with CCC 1.

There are at least two PEs in DCI strand ESS3, with CCC 2 and CCC 4.

The range of items per DCI strand is described as follows:

* Between one and two items aligned to PEs from DCI strand ESS1 will be assessed on Segment A of the CAST.
* Between one and five items aligned to PEs from DCI strand ESS2 will be assessed on Segment A of the CAST.
* Between one and three items aligned to PEs from DCI strand ESS3 will be assessed on Segment A of the CAST.

For the entire ESS domain, between eight and nine items will be assessed on Segment A of the CAST.

In the ETS subdomain for grade five, there are three PEs, organized into one DCI strand, that are distributed among three of the eight SEPs and no CCCs.

* For the science component of SEP 1, there is at least one PE.
* For SEP 3, there is at least one PE.
* For the engineering component of SEP 6 (SEP 6E), there is at least one PE.

**Alternative Text for Figure 4.B.1 *(continuation three)***

The range of items per DCI strand is described as follows:

* Between two and four items aligned to PEs from DCI strand ETS1 will be assessed on Segment A of the CAST.

For the entire ETS subdomain, between two and four items will be assessed on Segment A of the CAST.

The range of items per SEP across all domains in grade five is described as follows:

* Between one and four items representing both the science and engineering components of SEP 1 will be assessed on Segment A of the CAST.
* Between one and seven items representing SEP 2 will be assessed on Segment A of the CAST.
* Between one and seven items representing SEP 3 will be assessed on Segment A of the CAST.
* Between two and four items representing SEP 4 will be assessed on Segment A of the CAST.
* Between one and two items representing SEP 5 will be assessed on Segment A of the CAST.
* Between two and eight items representing both the science and engineering components of SEP 6 will be assessed on Segment A of the CAST.
* Between one and eight items representing SEP 7 will be assessed on Segment A of the CAST.
* Between one and three items representing SEP 8 will be assessed on Segment A of the CAST.

In grade five, a total of 26 items representing a selection of PEs across all three science domains and the ETS subdomain will be assessed on Segment A of the CAST.

*Return to figure 4.B.1.*

##### Alternative Text for Figure 4.B.2

In the table, an X indicates that there is at least one PE at the given intersection of the three dimensions that can be sampled on a test form for Segment A. The table has an X only in the locations described in the bulleted text that follows for each science domain and the ETS subdomain.

In the PS domain for grade eight, there are 19 PEs, organized into four DCI strands, that are distributed among eight SEPs and seven CCCs.

* For the science component of SEP 1, there is at least one PE in DCI strand PS2, with CCC 2.
* For SEP 2, there are at least five PEs.

There are at least three PEs in DCI strand PS1, with CCC 2, CCC 3, and CCC 5.

There is at least one PE in DCI strand PS3, with CCC 4.

There is at least one PE in DCI strand PS4, with CCC 6.

* For SEP 3, there are at least three PEs.

There are at least two PEs in DCI strand PS2, with CCC 2 and CCC 7.

There is at least one PE in DCI strand PS3, with CCC 3.

* For SEP 4, there are at least two PEs.

There is at least one PE in DCI strand PS1, with CCC 1.

There is at least one PE in DCI strand PS3, with CCC 3.

* For SEP 5, there is at least one PE in DCI strand PS4, with CCC 1.
* For the engineering component of SEP 6 (SEP 6E), there are at least three PEs.

There is at least one PE in DCI strand PS1, with CCC 5.

There is at least one PE in DCI strand PS2, with CCC 4.

There is at least one PE in DCI strand PS3, with CCC 5.

* For SEP 7, there are at least two PEs.

There is at least one PE in DCI strand PS2, with CCC 4.

There is at least one PE in DCI strand PS3, with CCC 5.

* For SEP 8, there are at least two PEs.

There is at least one PE in DCI strand PS1, with CCC 6.

There is at least one PE in DCI strand PS4, with CCC 6.

**Alternative Text for Figure 4.B.2 *(continuation one)***

The range of items per DCI strand is described as follows:

* Between one and five items aligned to PEs from DCI strand PS1 will be assessed on Segment A of the CAST.
* Between one and four items aligned to PEs from DCI strand PS2 will be assessed on Segment A of the CAST.
* Between one and four items aligned to PEs from DCI strand PS3 will be assessed on Segment A of the CAST.
* Between one and two items aligned to PEs from DCI strand PS4 will be assessed on Segment A of the CAST.

For the entire PS domain, between 8 and 10 items will be assessed on Segment A of the CAST.

In the LS domain for grade eight, there are 21 PEs, organized into four DCI strands, that are distributed among seven of the eight SEPs and seven CCCs.

* For SEP 2, there are at least five PEs.

There are at least two PEs in DCI strand LS1, with CCC 5 and CCC 6.

There is at least one PE in DCI strand LS2, with CCC 5.

There are at least two PEs in DCI strand LS3, with CCC 2 and CCC 6.

* For SEP 3, there is at least one PE in DCI strand LS1, with CCC 3.
* For SEP 4, there are at least two PEs.

There is at least one PE in DCI strand LS2, with CCC 2.

There is at least one PE in DCI strand LS4, with CCC 1.

* For SEP 5, there is at least one PE in DCI strand LS4, with CCC 2.
* For the science component of SEP 6 (SEP 6), there are at least five PEs.

There are at least two PEs in DCI strand LS1, with CCC 2 and CCC 5.

There is at least one PE in DCI strand LS2, with CCC 1.

There are at least two PEs in DCI strand LS4, with CCC 1 and CCC 2.

* For SEP 7, there are at least three PEs.

There are at least two PEs in DCI strand LS1, with CCC 2 and CCC 4.

There is at least one PE in DCI strand LS2, with CCC 7.

* For SEP 8, there are at least two PEs.

There is at least one PE in DCI strand LS1, with CCC 2.

There is at least one PE in DCI strand LS4, with CCC 2.

**Alternative Text for Figure 4.B.2 *(continuation two)***

The range of items per DCI strand is described as follows:

* Between one and six items aligned to PEs from DCI strand LS1 will be assessed on Segment A of the CAST.
* Between one and four items aligned to PEs from DCI strand LS2 will be assessed on Segment A of the CAST.
* Between one and two items aligned to PEs from DCI strand LS3 will be assessed on Segment A of the CAST.
* Between one and five items aligned to PEs from DCI strand LS4 will be assessed on Segment A of the CAST.

For the entire LS domain, between 8 and 10 items will be assessed on Segment A of the CAST.

In the ESS domain for grade eight, there are 15 PEs, organized into three DCI strands, that are distributed among six of the eight SEPs and six of the seven CCCs.

* For the science component of SEP 1, there is at least one PE in DCI strand ESS3, with CCC 7.
* For SEP 2, there are at least five PEs.

There are at least two PEs in DCI strand ESS1, with CCC 1 and CCC 4.

There are at least three PEs in DCI strand ESS2, with CCC 4, CCC 5, and CCC 7.

* For SEP 3, there is at least one PE in DCI strand ESS2, with CCC 2.
* For SEP 4, there are at least three PEs.

There is at least one PE in DCI strand ESS1, with CCC 3.

There is at least one PE in DCI strand ESS2, with CCC 1.

There is at least one PE in DCI strand ESS3, with CCC 1.

* For the science component of SEP 6, there are at least three PEs.

There is at least one PE in DCI strand ESS1, with CCC 3.

There is at least one PE in DCI strand ESS2, with CCC 3.

There is at least one PE in DCI strand ESS3, with CCC 2.

* For the engineering component of SEP 6 (SEP 6E), there is at least one PE in DCI strand ESS3, with CCC 2.
* For SEP 7, there is at least one PE in DCI strand ESS3, with CCC 2.

The range of items per DCI strand is described as follows:

* Between one and three items aligned to PEs from DCI strand ESS1 will be assessed on Segment A of the CAST.
* Between one and five items aligned to PEs from DCI strand ESS2 will be assessed on Segment A of the CAST.
* Between one and four items aligned to PEs from DCI strand ESS3 will be assessed on Segment A of the CAST.

For the entire ESS domain, between 8 and 10 items will be assessed on Segment A of the CAST.

**Alternative Text for Figure 4.B.2 *(continuation three)***

In the ETS subdomain for grade eight, there are four PEs, organized into one DCI strand, that are distributed among four of the eight SEPs and no CCCs.

* For the science component of SEP 1, there is at least one PE aligned to DCI strand ETS1.
* For SEP 2, there is at least one PE aligned to DCI strand ETS1.
* For SEP 4, there is at least one PE aligned to DCI strand ETS1.
* For SEP 7, there is at least one PE aligned to DCI strand ETS1.

The range of items per DCI strand is described as follows:

* Between two and four items aligned to PEs from DCI strand ETS1 will be assessed on Segment A of the CAST.

For the entire ETS subdomain, between two and four items will be assessed on Segment A of the CAST.

The range of items per SEP across all domains is described as follows:

* Between one and three items representing both the science and engineering components of SEP 1 will be assessed on Segment A of the CAST.
* Between 1 and 16 items representing SEP 2 will be assessed on Segment A of the CAST.
* Between one and five items representing SEP 3 will be assessed on Segment A of the CAST.
* Between one and nine items representing SEP 4 will be assessed on Segment A of the CAST.
* Between one and two items representing SEP 5 will be assessed on Segment A of the CAST.
* Between 1 and 12 items representing both the science and engineering components of SEP 6 will be assessed on Segment A of the CAST.
* Between one and eight items representing SEP 7 will be assessed on Segment A of the CAST.
* Between one and four items representing SEP 8 will be assessed on Segment A of the CAST.

For grade eight, a total of 28 items representing a selection of PEs across all three science domains and the ETS subdomain will be assessed on Segment A of the CAST.

*Return to figure 4.B.2.*

##### Alternative Text for Figure 4.B.3

In the table, an X indicates that there is at least one PE at the given intersection of the three dimensions that can be sampled on a test form for Segment A. The table has an X only in the locations described in the bulleted text that follows for each science domain and the ETS subdomain.

In the PS domain for high school, there are 24 PEs, organized into four DCI strands, that are distributed among the eight SEPs and six of the seven CCCs.

* For the science component of SEP 1, there is at least one PE in DCI strand PS4, with CCC 7.
* For SEP 2, there are at least four PEs.

There are at least two PEs in DCI strand PS1, with CCC 1 and CCC 5.

There are at least two PEs in DCI strand PS3, with CCC 2 and CCC 5.

* For SEP 3, there are at least three PEs.

There is at least one PE in DCI strand PS1, with CCC 1.

There is at least one PE in DCI strand PS2, with CCC 2.

There is at least one PE in DCI strand PS3, with CCC 4.

* For SEP 4, there is at least one PE in DCI strand PS2, with CCC 2.
* For SEP 5, there are at least five PEs.

There is at least one PE in DCI strand PS1, with CCC 5.

There are at least two PEs in DCI strand PS2, with CCC 1 and CCC 4.

There is at least one PE in DCI strand PS3, with CCC 4.

There is at least one PE in DCI strand PS4, with CCC 2.

* For the science component of SEP 6, there is at least one PE in DCI strand PS1, with CCC 1.
* For the engineering component of SEP 6 (SEP 6E), there are at least three PEs.

There is at least one PE in DCI strand PS1, with CCC 7.

There is at least one PE in DCI strand PS2, with CCC 2.

There is at least one PE in DCI strand PS3, with CCC 5.

* For SEP 7, there is at least one PE in DCI strand PS4, with CCC 4.
* For SEP 8, there are at least two PEs.

There is at least one PE in DCI strand PS2, with CCC 6.

There is at least one PE in DCI strand PS4, with CCC 2.

**Alternative Text for Figure 4.B.3 *(continuation one)***

The range of items per DCI strand is described as follows:

* Between two and seven items aligned to PEs from DCI strand PS1 will be assessed on Segment A of the CAST.
* Between one and five items aligned to PEs from DCI strand PS2 will be assessed on Segment A of the CAST.
* Between one and four items aligned to PEs from DCI strand PS3 will be assessed on Segment A of the CAST.
* Between one and four items aligned to PEs from DCI strand PS4 will be assessed on Segment A of the CAST.

For the entire PS domain, between 9 and 12 items will be assessed on Segment A of the CAST.

In the LS domain for high school, there are 24 PEs, organized into four DCI strands, that are distributed among eight SEPs and seven CCCs.

* For the science component of SEP 1, there is at least one PE in DCI strand LS3, with CCC 2.
* For SEP 2, there are at least three PEs.

There are at least two PEs in DCI strand LS1, with CCC 4 and CCC 5.

There is at least one PE in DCI strand LS2, with CCC 4.

* For SEP 3, there is at least one PE in DCI strand LS1, with CCC 7.
* For SEP 4, there are at least two PEs.

There is at least one PE in DCI strand LS3, with CCC 3.

There is at least one PE in DCI strand LS4, with CCC 1.

* For SEP 5, there are at least three PEs.

There are at least two PEs in DCI strand LS2, with CCC 3 and CCC 5.

There is at least one PE in DCI strand LS4, with CCC 2.

* For the science component of SEP 6, there are at least four PEs.

There are at least two PEs in DCI strand LS1, with CCC 5 and CCC 6.

There is at least one PE in DCI strand LS2, with CCC 5.

There is at least one PE in DCI strand LS4, with CCC 2.

* For the engineering component of SEP 6 (SEP 6E), there is at least one PE in DCI strand LS2, with CCC 7.

**Alternative Text for Figure 4.B.3 *(continuation two)***

* For SEP 7, there are at least four PEs.

There are at least two PEs in DCI strand LS2, with CCC 2 and CCC 7.

There is at least one PE in DCI strand LS3, with CCC 2.

There is at least one PE in DCI strand LS4, with CCC 2.

* For SEP 8, there is at least one PE in DCI strand LS4, with CCC 1.

The range of items per DCI strand is described as follows:

* Between one and six items aligned to PEs from DCI strand LS1 will be assessed on Segment A of the CAST.
* Between one and seven items aligned to PEs from DCI strand LS2 will be assessed on Segment A of the CAST.
* Between one and two items aligned to PEs from DCI strand LS3 will be assessed on Segment A of the CAST.
* Between one and five items aligned to PEs from DCI strand LS4 will be assessed on Segment A of the CAST.

For the entire LS domain, between 9 and 12 items will be assessed on Segment A of the CAST.

In the ESS domain for high school, there are 19 PEs, organized into three DCI strands, that are distributed among seven of the eight SEPs and seven CCCs.

* For SEP 2, there are at least four PEs.

There is at least one PE in DCI strand ESS1, with CCC 3.

There are at least three PEs in DCI strand ESS2, with CCC 2, CCC 5, and CCC 7.

* For SEP 3, there is at least one PE in DCI strand ESS2, with CCC 6.
* For SEP 4, there are at least two PEs.

There is at least one PE in DCI strand ESS2, with CCC 7.

There is at least one PE in DCI strand ESS3, with CCC 7.

* For SEP 5, there are at least three PEs.

There is at least one PE in DCI strand ESS1, with CCC 3.

There are at least two PEs in DCI strand ESS3, with CCC 4 and CCC 7.

* For the science component of SEP 6, there are at least three PEs.

There are at least two PEs in DCI strand ESS1, with CCC 5 and CCC 7.

There is at least one PE in DCI strand ESS3, with CCC 2.

* For the engineering component of SEP 6 (SEP 6E), there is at least one PE in DCI strand ESS3, with CCC 7.

**Alternative Text for Figure 4.B.3 *(continuation three)***

* For SEP 7, there are at least three PEs.

There is at least one PE in DCI strand ESS1, with CCC 1.

There is at least one PE in DCI strand ESS2, with CCC 7.

There is at least one PE in DCI strand ESS3, with no CCC.

* For SEP 8, there is at least one PE in DCI strand ESS1, with CCC 5.

The range of items per DCI strand is described as follows:

* Between one and five items aligned to PEs from DCI strand ESS1 will be assessed on Segment A of the CAST.
* Between one and six items aligned to PEs from DCI strand ESS2 will be assessed on Segment A of the CAST.
* Between one and five items aligned to PEs from DCI strand ESS3 will be assessed on Segment A of the CAST.

For the entire ESS domain, between 9 and 12 items will be assessed on Segment A of the CAST.

In the ETS subdomain for high school, there are four PEs, organized into one DCI strand, that are distributed among three of the eight SEPs and one of the seven CCCs.

* For the engineering component of SEP 1 (SEP 1E), there is at least one PE in the DCI strand ETS1, with no CCC.
* For SEP 5, there is at least one PE in the DCI strand ETS1, with CCC 4.
* For the science component of SEP 6, there is at least one PE in the DCI strand ETS1, with no CCC.

The range of items per DCI strand is described as follows:

* Between two and four items aligned to PEs from DCI strand ETS1 will be assessed on Segment A of the CAST.

For the entire ETS subdomain, between two and four items will be assessed on Segment A of the CAST.

The range of items per SEP across all domains is described as follows:

* Between two and three items representing both the science and engineering components of SEP 1 will be assessed on Segment A of the CAST.
* Between two and six items representing SEP 2 will be assessed on Segment A of the CAST.
* Between two and five items representing SEP 3 will be assessed on Segment A of the CAST.
* Between two and five items representing SEP 4 will be assessed on Segment A of the CAST.
* Between two and six items representing SEP 5 will be assessed on Segment A of the CAST.

**Alternative Text for Figure 4.B.3 *(continuation four)***

* Between two and six items representing both the science and engineering components of SEP 6 will be assessed on Segment A of the CAST.
* Between two and six items representing SEP 7 will be assessed on Segment A of the CAST.
* Between two and six items representing SEP 8 will be assessed on Segment A of the CAST.

For high school, a total of 32 items representing a selection of PEs across all three science domains and the ETS subdomain will be assessed on Segment A of the CAST.

*Return to figure 4.B.3.*

## Test Administration

This chapter details the processes involved in the administration of the 2023–24 California Science Test (CAST). It also describes the procedures followed by ETS to maintain test security throughout the test administration process.

### Overview

The CAST was administered to students in grade five, grade eight, and the high school grade band (grade ten, eleven, or twelve) in 2023–24 in conjunction with the other assessments that compose the California Assessment of Student Performance and Progress (CAASPP) System.

In accordance with the procedures for the computer-based CAASPP, local educational agencies (LEAs) identified test administrators and entered the test administrators as users into the Test Operations Management System (TOMS). ETS provided LEA staff with the appropriate training materials, such as test administration manuals, videos, and webinars, to ensure that the LEA staff and test administrators understood how to administer the computer-based CAST content-area assessments.

The testing window for the 2023–24 administration of CAST was January 9 through June 28, 2024. Specific test administration schedules within that window were determined locally pursuant to *California Code of Regulations*,Title 5 (5*CCR*), Section 855(a).

#### In-Person and Remote Testing

Table 5.1 shows the number of students who took the CAST in each grade level, the high school grade band, and the sum of all grade levels, broken down by test administration type (in person or remote). The percentage of students who tested in person ranged from 95.61 percent to 97.97 percent across all grade levels and the high school grade band.

Table 5.1 Number of Students by Assessment Location

|  |  |  |
| --- | --- | --- |
| **Grade Level or Grade Band** | **In-Person Assessment** | **Remote Assessment** |
| Grade 5 | 413,285 | 8,991 |
| Grade 8 | 416,134 | 11,169 |
| High school—Grade 10 | 23,832 | 1,094 |
| High school—Grade 11 | 323,245 | 8,631 |
| High school—Grade 12 | 97,549 | 2,019 |
| High school—All grades | 444,626 | 11,744 |
| **Totals:** | **1,274,045** | **31,904** |

#### Student Test-Taking Requirements

CAST was administered to students in grades five and eight as well as high school students in grade ten, eleven, or twelve who were not repeating grade twelve and who were assigned by their LEA. CAST is a science assessment for the general student population (that is, those students who are not otherwise eligible for the California Alternate Assessment [CAA] for Science). Subsection [*5.1.2.2 High School*](#_High_School) outlines the process for grade-level assignment for the CAST for high school students.

##### Grades Five and Eight

All students enrolled in grades five and eight were automatically designated in TOMS to take the CAST. If the student’s individualized education program (IEP) team indicated an alternate assessment, the LEA and school had to register the eligible student manually in TOMS to take the CAA for Science.

##### High School

At the high school level, schools and LEAs were responsible for assigning students in grade ten or eleven to take the CAST or CAA for Science. Guidelines were provided by the California Department of Education (CDE) suggesting that students who completed or were in the process of completing their last high school science course should take the science assessment, either the CAST or CAA for Science, depending on the student’s eligibility. All grade twelve students who had not previously completed the CAST in grade ten or eleven were automatically assigned to take the CAST. Neither students in grade nine nor students who repeated grade twelve were eligible to take a science assessment (CDE, 2021g).

### User Roles and Standardization

The test administration procedures were designed so that the assessments are administered in a standardized manner. ETS took all necessary measures to ensure the standardization of test administration, as described in this section.

#### Local Educational Agency CAASPP Coordinator

An LEA CAASPP coordinator was designated by the district superintendent or charter school administrator at the beginning of the 2023–24 school year. LEAs include public school districts, California State Board of Education–authorized charter schools, county office of education programs, and direct funded charter schools.

LEA CAASPP coordinators were responsible for ensuring the proper and consistent administration of the CAASPP. In addition to the responsibilities set forth in 5*CCR* Section 857, their responsibilities included

* adding site CAASPP coordinators and test administrators into TOMS;
* training site CAASPP coordinators and test administrators regarding the state requirements and CAASPP administration as well as security policies and procedures;
* providing checklists for site CAASPP coordinators and test administrators to review in preparation for administering the summative assessments;
* overseeing test administration activities;
* reporting test security incidents (including testing irregularities) to the CDE using the online Security and Test Administration Incident Reporting System (STAIRS)/Appeals process; and
* requesting an Appeal (if indicated by TOMS prompts while reporting an incident using the STAIRS/Appeals process).

#### Site CAASPP Coordinator

A site CAASPP coordinator is trained by the LEA CAASPP coordinator for each test site (5*CCR* Section 857[f]). A site CAASPP coordinator must be an employee of the LEA and must sign a security agreement (5 *CCR* Section 859[a]).

A test site coordinator was responsible for identifying test administrators and ensuring that they have signed CAASPP *Test Security Affidavits* (5 *CCR* Section 859[d]). A site CAASPP coordinator’s duties may have included

* adding test administrators into TOMS;
* entering test settings for students;
* creating testing schedules and procedures for a school consistent with state and LEA policies;
* working with technology staff to ensure secure browsers are installed and any technical issues are resolved;
* monitoring testing progress during the testing window and ensuring students take the CAST, as appropriate;
* coordinating and verifying the correction of student data errors in the California Longitudinal Pupil Achievement Data System;
* ensuring a student’s test session is rescheduled, if necessary;
* addressing testing problems;
* reporting test security incidents (including testing irregularities) to the CDE using the online STAIRS/Appeals process;
* overseeing administration activities at a school site; and
* requesting an Appeal (if indicated by TOMS prompts while reporting an incident using the STAIRS/Appeals process).

#### Test Administrator

Test administrators were identified by site CAASPP coordinators as individuals who would administer CAST.

A test administrator must have signed a security affidavit (5 *CCR* Section 859[d]).

A test administrator’s duties may have included

* ensuring the physical conditions of the testing room meet the criteria for a secure test environment;
* administering the CAASPP, including CAST;
* reporting all test security incidents to the site CAASPP coordinator and LEA CAASPP coordinator in a manner consistent with state and LEA policies;
* viewing student information prior to testing to ensure that the correct student receives the proper assessment with appropriate resources and reporting potential data errors to site CAASPP coordinators and LEA CAASPP coordinators;
* monitoring student progress throughout the test session using the Test Administrator Interface; and
* fully complying with all directions provided in the *Directions for Administration* (*DFAs*) for CAST (CDE, 2024d).

#### Instructions for Test Administration

##### *Directions for Administration*

The *DFAs*, used by test administrators to administer the CAST to students, are included in the *CAASPP Online Test Administration Manual* (CDE, 2024d). Test administrators must follow all directions and guidelines and read, word-for-word, the instructions to students in the “SAY” boxes to ensure standardization of test administration. Additionally, the *CAASPP Online Test Administration Manual* provided information to test administrators regarding the systems involved in testing, including sections describing the Test Delivery System (TDS), so test administrators could become familiar with the testing application used by their students (CDE, 2024d).

##### *CAASPP Online Test Administration Manual*

The *CAASPP Online Test Administration Manual* (CDE, 2024d) contained information and instructions on overall procedures and guidelines for all LEA and test site staff involved in the administration of computer-based assessments. Sections included the following topics:

* Test administration resources
* Test security
* Administration preparation and planning
* General test administration
* In-person test administration
* Remote test administration
* Test administration directions and scripts for test administrators
* Overview of the student testing application
* Instructions for steps to take before, during, and after testing

Appendices included definitions of common terms and descriptions of different aspects of the assessment and systems associated with the assessment.

##### *Test Operations Management System User Guide*

TOMS is a web-based application that allows LEA CAASPP coordinators to set up test administrations, add and manage users, submit computer-based student test settings, and order paper–pencil tests (PPTs).

TOMS modules described in the *TOMS User Guide* included the following (CDE, 2024e):

* **Test Administration Setup—**This module allowed LEAs to determine and calculate dates for the LEA 2023–24 administration of the CAASPP, including CAST.
* **Adding and Managing Users—**This module allowed LEA CAASPP coordinators to add site CAASPP coordinators and test administrators to TOMS so that the designated user could administer, monitor, and manage the CAASPP computer-based assessments. The manual contained descriptions of the roles and responsibilities of those involved with CAASPP testing.
* **Reports—**This module allowed LEA CAASPP coordinators and site CAASPP coordinators access to the various reports in TOMS.
* **STAIRS/Appeals—**This module allowed LEA CAASPP coordinators and site CAASPP coordinators access to create new STAIRS cases or search for STAIRS/Appeals cases.
* **Student Profile—**This module allowed LEA CAASPP coordinators, site CAASPP coordinators, and test administrators and test examiners to view and manage student’s test assignments and test settings.

##### Other System Manuals

Other manuals were created to assist LEA CAASPP coordinators and others with the technological components of the CAASPP System and are listed next.

* ***CAASPP and English Language Proficiency Assessments for California (ELPAC) Technical Specifications and Configuration Guide for Online Testing*—**This manual provided information, tools, and recommended configuration details to help technology staff prepare computers and install the secure browser to be used for the computer-based CAASPP (CDE, 2024c).
* ***CAASPP and ELPAC Security Incidents and Appeals Procedure Guide*—**This manual provided information on how to report a testing incident and submit an Appeal to reset, reopen, invalidate, or restore individual computer-based student assessments (CDE, 2024b).
* ***CAASPP and ELPAC Accessibility Guide*—**This manual provided descriptions of the accessibility features for computer-based assessments as well as information about supported hardware and software requirements for administering assessments to students using accessibility resources, including those with a braille accommodation using Job Access With Speech® (software) or a braille embosser (hardware) (CDE, 2024a).

### Local Educational Agency Training

Each year, ETS, in collaboration with the CDE and its Assessment Validity and Outreach contractor, the Sacramento County Office of Education (SCOE), establishes and implements a comprehensive training plan for LEA assessment staff and educators on all aspects of the assessment program. The ETS and SCOE annual training plans specify the audience, topics, frequency, and mode (synchronous or asynchronous) of the training, including such elements as format, participants, and organization.

ETS and SCOE make every effort to make the information available in a variety of ways that allowed educators flexibility in accessing training at a time that best fits in with their day-to-day activities. This includes offering training events on multiple days and times, livestreaming events, recording and archiving training, and converting training to self-paced modules that can be taken any time, at the learner’s convenience.

All training opportunities are posted in one centralized location on the CAASPP & ELPAC Website. LEA staff can register for training opportunities in one place, on the Upcoming and On-Demand Trainings web page. Archived training is also made available on this web page, making it easier for educators to find a training they missed and providing easy access to recorded trainings. Participants can also register to receive a copy of the training materials without registering to attend the live training. Training materials are developed in such a way that educators can consume the information independently by accessing the training materials.

#### Synchronous and Asynchronous Training

All synchronous training for 2023–24 was offered on Zoom, recorded, and made available for on-demand viewing. Zoom provides an opportunity for educators to ask questions and get answers in real time. Coffee Sessions were live streamed on YouTube.

ETS and SCOE use various strategies to increase engagement during synchronous trainings. Live polls are presented to get real-time feedback about attendees’ knowledge of a particular topic, allowing presenters to tailor presentations to the audience’s level of understanding. The chat functionality is enabled to give participants an opportunity to interact with each other or to provide open-ended feedback, or it is disabled to minimize distraction and drive attendees’ focus to the information being presented. Breakout groups are used in smaller group trainings, as appropriate. Breaks and processing time are incorporated into presentations to give attendees opportunities to attend to other responsibilities that could otherwise distract from the training.

Working closely with the CDE, ETS and SCOE provide informal support to educators by offering monthly Coffee Sessions. Coffee Sessions include CDE and ETS staff who can answer questions about all aspects of testing. ETS also offers several Office Hours for coordinators where support staff are generally available from 9 a.m. to 3 p.m., allowing coordinators to join as needed and get customized support. SCOE continues to offer Assessment and Accountability Information Meetings intended to provide LEA coordinators with regular updates about California’s assessment and accountability systems. All trainings and meetings are recorded and archived for on-demand viewing on the Upcoming and On-Demand Trainings web page on the CAASPP & ELPAC Website.

#### Videos and Guides

ETS produced videos on various aspects of administering the CAASPP, including how to perform functions within TOMS, such as setting up a test administration window, adding users, assigning assessments to students, and uploading test settings. SCOE produced the accompanying quick reference guides, providing multiple avenues of support for educators administering the assessments.

In addition to the standard administration videos, ETS produced additional videos to support administration. Some videos were geared toward parents/guardians to help them understand the Student Score Reports (SSRs). Other videos were intended to help coordinators or other users complete a process, such as how to administer an assessment from start to finish, administering a practice or training test, starting and stopping a test session, how to monitor student completion, and how to complete second scoring that is required for some of the assessments. This list is a sampling of the available videos intended to capture the major areas of support for various interest holders. The comprehensive suite of training videos can be found on the CAASPP & ELPAC Website.

#### Training for Proper Identification and Assignment of Designated Supports and Accommodations

ETS produced and updated additional short demonstration videos to maintain the library of videos for every embedded accessibility resource. The videos demonstrate how to use the resource for educators, students, and parents/guardians and were available in both English and Spanish on the Accessibility Resource Demonstration Videos web page on the CAASPP & ELPAC Website. Demonstration videos for the most frequently used non-embedded accessibility resources are also available. These videos were linked within the Individual Student Assessment Accessibility Profile (ISAAP) Tool, increasing access to the demonstration videos. Educators using the ISAAP Tool to determine the student’s needs could view the corresponding demonstration video without having to navigate away from the tool.

A video on how to use the ISAAP Tool was also available to support educators in the process of creating an individual student profile and matching accessibility resources to student needs to ensure a fair and valid testing experience for all students.

For the 2023–24 CAASPP administration, ETS produced a two-part asynchronous training module. Module A, Matching Accessibility Resources to Students’ Needs, focused on providing educators with an understanding of the importance of accessibility resources, the categories of accessibility resources, and the process for matching students with appropriate accessibility resources for daily instruction and on assessments. Module B, Using Accessibility Resources in Daily Instruction, focused on the importance of removing barriers to student learning and using accessibility resources in daily instruction. Educators could complete the training independently or had the option to attend one of two live sessions held by ETS to extend and deepen the learning experience.

At the California Assessment Conference, SCOE offered four sessions on accessibility:

1. “Building Bridges to Accessibility: Creating a Systematic Approach for Successful Implementation of Accessibility Resources” focused on approaches to implementing accessibility resources in instruction and assessment.
2. “From Awareness to Action: Implementing Accessibility Resources in the Classroom” aimed to bridge the gap between awareness and action, sharing knowledge and strategies for creating inclusive learning environments; its focus was supporting students with disabilities and English learner students.
3. “Advancing Equity Through Assessment Accessibility” explored using Universal Design for Learning principles to meet student needs.
4. The “Accessibility Tools for Matching Resources to students” focused on the resources available to students participating in CAASPP and provided information on matching student needs and using those resources.

#### Feedback for Continuous Improvement Survey

The CAASPP and ELPAC programs solicit feedback annually from various interest holder groups, including LEA CAASPP coordinators, LEA ELPAC coordinators, site CAASPP coordinators, site ELPAC coordinators, test administrators, test examiners, and users with the IA Administrator Only (used to administer only interim assessments) role through the CAASPP and ELPAC Feedback for Continuous Improvement Survey. Feedback was collected via a post-test survey sent to more than 293,000 California educators and through focus groups. Educators provided valuable feedback for potential improvements to the future administration of CAASPP and the ELPAC—one or both—by reporting some lessons they learned in 2023–24. This year, additional efforts were made to collect information about the newly released CAST Interim Assessments and ELPAC Interim Assessments.

Improvements made in response to survey results are detailed in [chapter 11](#_Continuous_and_Systematic). The CDE and ETS used key recommendations from educators to implement positive changes in the next test administration year.

##### Overview

The California educators who responded provided specific, actionable insights about their test administration experience. This survey gathered information and data from educators who were part of the administration of CAASPP, the ELPAC, or both programs. Its goal was to highlight successes and identify areas for improvement, both immediate and long term.

Overall, California educators continue to express positive experiences in their preparations for administering CAASPP and the ELPAC.

##### Communication

During the 2023–24 test administration year, the CDE and ETS continued to streamline communications and provide LEAs with relevant information throughout the year. CAASPP and ELPAC monthly communications were sent throughout the administration with timely reminders and training announcements. In addition, proactive communications were sent to help remind LEA CAASPP coordinators of important actions needed for a successful administration, such as reminders to set up a test administration window, order materials, or enter scores into the Data Entry Interface, if needed.

### Accessibility Resources

The Every Student Succeeds Act reaffirms the importance of ensuring that assessments are accessible to special student populations, and the Individuals with Disabilities Education Act lays out monitoring requirements for students with disabilities. This section describes the accessibility resources used to support students in the CAST, as well as the procedures to identify and assign students with accommodations and designated supports. Finally, the number of students who were assigned accessibility resources is reported on the basis of available data.

The 2023–24 CAST offered commonly used accessibility resources available through the CAASPP computer-based testing platform, where applicable for the tested construct.

#### Accessibility Resource Categories

The purpose of universal tools, designated supports, and accommodations in testing is to provide *all* students with the opportunity to demonstrate what they know and what they are able to do. Universal tools, designated supports, and accommodations minimize or remove barriers that could otherwise prevent students from demonstrating their knowledge, skills, and achievement in a specific content area.

The CDE California Assessment Accessibility Resources Matrix (Accessibility Matrix) (CDE, 2023) is intended for school-level personnel and IEP and Section 504 plan teams to select and administer the appropriate universal tools, designated supports, and accommodations as deemed necessary for individual students.

##### Universal Tools

Universal tools were available to all students by default, although they could be disabled if a student found them distracting. Each universal tool fell into one of two categories: embedded and non-embedded. Embedded universal tools were provided through the TDS (through the CAASPP secure browser), although they could be turned off by a test administrator.

The universal tools in the following subsections were available in the 2023–24 CAST administration.

###### Embedded

The following embedded universal tools were available to students testing in the secure browser:

* Breaks
* Calculator:[[3]](#footnote-4)

Four-function—grade five

Scientific—grade eight and high school

* Digital notepad
* English glossary
* Expandable items[[4]](#footnote-5)
* Expandable passages
* Highlighter (embedded)
* Keyboard navigation
* Line reader (embedded)
* Mark for review
* Math tools (for example, ruler, protractor)[[5]](#footnote-6)
* Science charts (that is, periodic table of the elements, reference sheet)
* Science tools (for example, analog clock, laboratory equipment)
* Spell check
* Strikethrough
* Writing tools (for example, bold, italic, bullets, undo/redo)
* Zoom (in/out)

###### Non-Embedded

The following non-embedded universal tools were available to students testing in the secure browser:

* Breaks
* Scratch paper

The following non-embedded universal tools were available to students taking the PPT:

* Breaks
* Calculator
* English glossary
* Highlighter
* Line reader
* Mark for review
* Science reference tools
* Scratch paper
* Strikethrough

##### Designated Supports

Designated supports were available to all students when determined for use by an educator or team of educators (with parent/guardian and student input, as appropriate) or specified in the student’s IEP or Section 504 plan. These are assigned through the test settings in TOMS. The designated supports each fell into one of two categories: embedded and non-embedded. Embedded designated supports were provided through the Student Testing Interface (through the CAASPP secure browser).

The designated supports in the following subsections were available in the 2023–24 CAST administration.

###### Embedded

The following embedded designated supports were available to students testing in the secure browser:

* Color contrast
* Masking
* Mouse pointer (size and color)
* Permissive mode
* Print (font) size
* Streamline
* Text-to-speech (items and stimuli)
* Translated test directions
* Translations (glossary)
* Translations (Spanish stacked–dual language)
* Turn off any universal tool(s)

###### Non-Embedded

The following non-embedded designated supports were available to students testing in the secure browser:

* 100s number table
* Amplification
* Calculator:

Four-function—grade five

Scientific—grade eight and high school

* Color contrast
* Color overlay
* Magnification
* Medical supports
* Multiplication table
* Noise buffers
* Read aloud (items and stimuli)
* Read aloud in Spanish
* Science charts (that is, periodic table of the elements, reference sheet)[[6]](#footnote-7)
* Scribe
* Separate setting (for example, most beneficial time, special lighting or acoustics, adaptive furniture)
* Simplified test directions
* Translated test directions

The following non-embedded designated supports were available to students taking the PPT:

* 100s number table
* Color overlay
* Magnification
* Masking
* Medical supports
* Multiplication table
* Noise buffers
* Read aloud (items)
* Science charts (that is, periodic table of the elements, reference sheet)
* Scribe
* Separate setting (for example, most beneficial time, special lighting or acoustics, adaptive furniture)
* Simplified test directions
* Translated test directions
* Translations (glossary)

##### Accommodations

Accommodations are changes in procedures or materials that increased equitable access during CAASPP administration and are permitted to all eligible students if specified in the student’s IEP or Section 504 plan. Assessment accommodations for students who needed them generated valid assessment results; they allowed these students to show what they know and can do. Accommodations did not compromise the learning expectations, construct, grade-level standard, or intended outcome of the assessments.

The accommodations in the following subsections were available in the 2023–24 CAST administration.

###### Embedded

The following embedded accommodations were available to students testing in the secure browser:

* American Sign Language (videos)
* Braille (embosser and refreshable)
* Closed-captioning (allowed but not currently used)
* Speech-to-text
* Word prediction

###### Non-Embedded

The following non-embedded accommodations were available to students testing in the secure browser:

* Abacus
* Alternate response options
* Print-on-demand
* Speech-to-text
* Word prediction

The following non-embedded accommodations were available to students taking the PPT:

* Abacus
* Alternate response options
* American Sign Language
* Braille
* Breaks
* Large print

##### Unlisted Resources

An unlisted resource is an instructional support a student regularly uses in daily instruction, assessment, or both, and has not been previously identified as a universal tool, designated support, or accommodation. The CDE Accessibility Matrix included an inventory of unlisted resources that were already identified and were preapproved (CDE, 2023). During the 2023–24 CAASPP administration, an LEA CAASPP coordinator or a site CAASPP coordinator would use TOMS to submit a request for use of an unlisted resource. A preidentified, preapproved unlisted resource request was automatically approved. A request for an unlisted resource that was not preidentified was sent to the CDE for review and adjudication.

Unlisted resources are non-embedded resources that are made available if specified in the eligible student’s IEP or Section 504 plan and only upon approval by the CDE. Unlisted resources that changed the construct of an assessment and were approved were flagged as causing a change in construct. Test results for a student using an unlisted resource that was approved but that changed the construct of what was being tested were considered invalid for reporting purposes.

Preidentified unlisted resources applicable to CAST are as follows:

* Bilingual dictionary
* English dictionary
* Signed exact English
* Thesaurus
* Translated word lists
* Translations (not provided by the CDE, Smarter Balanced, or ETS)

The LEA CAASPP coordinator or site CAASPP coordinator was required to submit a request for the use of an unlisted resource to the CDE a minimum of 10 business days before the student’s first day of testing.

#### Identification and Selection

All eligible students enrolled in a California public school participate in the CAASPP System, including students with disabilities and English learner students. The Smarter Balanced Assessment Consortium’s *Usability, Accessibility, and Accommodations Guidelines* (Smarter Balanced, 2023) and the CDE Accessibility Matrix (CDE, 2023) are intended for school-level personnel and IEP and Section 504 plan teams to select and administer the appropriate universal tools, designated supports, and accommodations as deemed necessary for individual students.[[7]](#footnote-8) CAST follows the Smarter Balanced recommendations for use (Smarter Balanced, 2018).

The *Guidelines* apply to all participating students and promote an individualized approach to the implementation of assessment practices. Another web page, the Smarter Balanced Accessibility Strategies web page on the Tools for Teachers website (Smarter Balanced, 2024), connects the assessment resources described in the *Guidelines* with associated classroom practices.

The full list of the universal tools, designated supports, and accommodations used in CAASPP computer-based assessments, including CAST, is documented in the CDE Accessibility Matrix. Most embedded and non-embedded universal tools, designated supports, and accommodations listed in parts 1, 2, and 3 of the Accessibility Matrix are available for CAST through the computer-based testing interface or, in the case of non-embedded resources, from the school or LEA. Part 5 of the Accessibility Matrix includes approved unlisted resources. School-level personnel, IEP teams, and Section 504 teams used the CDE Accessibility Matrix when deciding how best to support the student’s test-taking experience. Another manual, the *Smarter Balanced Usability, Accessibility, and Accommodations Implementation Guide* (Smarter Balanced, 2014),provides suggestions for implementation of these resources.

Test administrators and test examiners are given the opportunity to administer the CAASPP practice and training tests so that students have the opportunity to familiarize themselves with a designated support or accommodation prior to testing. (Refer to section [*5.5 Practice and Training Tests*](#_Practice_and_Training) for additional information.)

#### Assignment

Designated supports and accommodations are assigned to individual students on the basis of identified student need. Such assignments are implemented in TOMS by the LEA CAASPP coordinator or site CAASPP coordinator, either through individual assignment through the student’s profile in TOMS or in a batch upload for multiple students. When the batch upload process was used, settings were uploaded into TOMS using a spreadsheet with data that had either been entered into a template downloaded from TOMS; or created by selecting and entering information into the web-based ISAAP Tool. The ISAAP Tool could be used by LEAs in conjunction with the *Guidelines* and the *2023–24* CAASPP and ELPAC Accessibility Guide (CDE, 2024a), as well as with state regulations and policies (such as the Accessibility Matrix) related to assessment accessibility*.*

The embedded designated supports and accommodations were delivered to the student through the TDS at the time of testing; the non-embedded designated supports and accommodations were provided at the time of testing to the student by the LEA. Refer to section [*1.9 Systems Overview and Functionality*](#_Systems_Overview_and) in [*Chapter 1: Introduction*](#_Introduction) for more details regarding the TDS.

Once a student’s IEP or Section 504 plan team decided which accessibility resource(s) the student should use, LEA CAASPP coordinators and site CAASPP coordinators used TOMS to assign designated supports and accommodations to students prior to the start of a test session.

There were three ways a student’s accessibility resource(s) could be assigned:

1. Using the ISAAP Tool to identify the accessibility resource(s) and then uploading the spreadsheet it creates into TOMS (This process is discussed in more detail in subsection [*5.4.2 Identification and Selection*](#_Identification_and_Selection).)
2. Using the Online Student Test Settings template to enter students’ assignments and then uploading the spreadsheet into TOMS
3. Entering assignments for each student individually in TOMS

If a student’s IEP or Section 504 plan team identified and designated a resource not identified in the CDE Accessibility Matrix, the LEA CAASPP coordinator or site CAASPP coordinator needed to submit a request for an unlisted resource to be approved by the CDE. The CDE then determined whether the requested unlisted resource changed the construct being measured before the student started testing. Once the requested unlisted resource was approved, the student could begin the assessment using the approved resource.

[Appendix 5.A](#_Appendix_5.A:_Accessibility_1) provides information on the number of students who were assigned accommodations and designated supports.

#### Delivery of Embedded and Non-Embedded Resources to Students

Universal tools, designated supports, and accommodations can be delivered as either embedded or non-embedded resources. Embedded resources are digitally delivered features or settings available as part of the technology platform for CAST testing. Examples of embedded resources include the expandable items, color contrast, and masking.

Non-embedded resources are available, when provided by the LEA, for both computer-based assessments and PPTs. These resources are not part of the technology platform for the computer-administered CAST. Examples of non-embedded resources include magnification, noise buffers, and the use of a scribe.

Refer to subsection [*5.4.1 Accessibility Resource Categories*](#_Accessibility_Resource_Categories_1) for a detailed description of the accessibility resources available to students taking CAST.

#### Usage of Designated Supports and Accommodations

LEA CAASPP coordinators and site CAASPP coordinators were responsible for assigning their students’ test settings in TOMS before testing occurred and providing the necessary resources during testing. If a test setting was not applied before testing, then a STAIRS incident was to be submitted to reset the assessment so the student could be retested with the correct accommodation or designated support. If a test setting was accidentally assigned to a student, then a STAIRS incident was also to be submitted to reset the assessment so the student could be retested without the accommodation or designated support.

After schools and LEAs assigned eligible students to accommodations or designated supports, the Cambium Assessment, Inc. (CAI) TDS provided and captured whether a certain accommodation or designated support (or multiple accommodations or designated supports) was used by a student—that is, the student interacted with a control for the resource, such as a button—as the student progressed through the assessment.

[Appendix 5.A](#_Appendix_5.A:_Accessibility_1) reports the number of students who were assigned to a certain accommodation or designated support at school. Table 5.2 reports the number of students who, on the basis of the availability of data, were assigned to each accommodation or designated support and the number of students who used this accommodation or designated support at least once during test administration. However, because the TDS does not capture the usage of all embedded resources and cannot capture the usage of any non-embedded resources, this table reports only on a limited subset of the embedded resources. (Embedded accessibility resources are those that are part of the computer-based TDS, whereas non-embedded accessibility resources are provided outside of that system.)

Types of accommodations and designated supports—labeled “ACC” and “DS” in the *Resource Type* column—included in table 5.2 are as follows:

* **American Sign Language (ASL):** ASL videos are available for any item that has a listening component. The ASL human signer and the signed test content are viewed on the same screen.
* **Speech-to-Text:** Voice recognition allows students to use their voices as input devices to the computer to dictate responses.
* **Print-on-Demand:** Paper copies of passages and stimuli, items, or all of these are printed for students.
* **Masking:** This resource involves blocking off content that is not of immediate need or that may be distracting to the student.
* **Text-to-Speech (TTS):** Text is read aloud to the student via embedded TTS technology either in English or used in conjunction with the translations (Spanish stacked–dual language) designated support.

Table 5.2 Summary of Accommodations and Designated Supports Used by Students

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Accessibility Resource** | **Resource Type** | **Students Assigned** | **Students Used** | **Percentage Used** |
| All | Any Tracked Resource | All | 161,553 | 60,204 | 37.27 |
| All | Embedded American Sign Language | ACC | 445 | 248 | 55.73 |
| All | Embedded Speech-to-Text | ACC | 28,420 | 8,930 | 31.42 |
| All | Non-Embedded Print-on-Demand | ACC | 286 | 39 | 13.64 |
| All | Embedded Masking | DS | 32,829 | 1,461 | 4.45 |
| All | Embedded Text-to-Speech (English TTS) | DS | 141,340 | 55,631 | 39.36 |
| Grade 5 | Any Tracked Resource | All | 75,626 | 41,616 | 55.03 |
| Grade 5 | Embedded American Sign Language | ACC | 159 | 108 | 67.92 |
| Grade 5 | Embedded Speech-to-Text | ACC | 16,226 | 7,379 | 45.48 |
| Grade 5 | Non-Embedded Print-on-Demand | ACC | 114 | 27 | 23.68 |
| Grade 5 | Embedded Masking | DS | 11,933 | 636 | 5.33 |
| Grade 5 | Embedded Text-to-Speech (English TTS) | DS | 67,034 | 38,549 | 57.51 |
| Grade 8 | Any Tracked Resource | All | 53,121 | 14,565 | 27.42 |
| Grade 8 | Embedded American Sign Language | ACC | 149 | 72 | 48.32 |
| Grade 8 | Embedded Speech-to-Text | ACC | 8,842 | 1,382 | 15.63 |
| Grade 8 | Non-Embedded Print-on-Demand | ACC | 67 | 6 | 8.96 |
| Grade 8 | Embedded Masking | DS | 9,967 | 494 | 4.96 |
| Grade 8 | Embedded Text-to-Speech (English TTS) | DS | 46,824 | 13,507 | 28.85 |
| High school—Grade 10 | Any Tracked Resource | All | 1,978 | 352 | 17.80 |
| High school—Grade 10 | Embedded American Sign Language | ACC | 12 | 9 | 75.00 |
| High school—Grade 10 | Embedded Speech-to-Text | ACC | 252 | 12 | 4.76 |
| High school—Grade 10 | Non-Embedded Print-on-Demand | ACC | 1 | 0 | 0.00 |
| High school—Grade 10 | Embedded Masking | DS | 273 | 9 | 3.30 |
| High school—Grade 10 | Embedded Text-to-Speech (English TTS) | DS | 1,799 | 329 | 18.29 |
| High school—Grade 11 | Any Tracked Resource | All | 25,933 | 3,160 | 12.19 |
| High school—Grade 11 | Embedded American Sign Language | ACC | 92 | 38 | 41.30 |
| High school—Grade 11 | Embedded Speech-to-Text | ACC | 2,620 | 137 | 5.23 |
| High school—Grade 11 | Non-Embedded Print-on-Demand | ACC | 74 | 3 | 4.05 |
| High school—Grade 11 | Embedded Masking | DS | 8,813 | 273 | 3.10 |
| High school—Grade 11 | Embedded Text-to-Speech (English TTS) | DS | 21,367 | 2,812 | 13.16 |
| High school—Grade 12 | Any Tracked Resource | All | 4,895 | 511 | 10.44 |
| High school—Grade 12 | Embedded American Sign Language | ACC | 33 | 21 | 63.64 |
| High school—Grade 12 | Embedded Speech-to-Text | ACC | 480 | 20 | 4.17 |
| High school—Grade 12 | Non-Embedded Print-on-Demand | ACC | 30 | 3 | 10.00 |
| High school—Grade 12 | Embedded Masking | DS | 1,843 | 49 | 2.66 |
| High school—Grade 12 | Embedded Text-to-Speech (English TTS) | DS | 4,316 | 434 | 10.06 |
| High school—All grades | Any Tracked Resource | All | 32,806 | 4,023 | 12.26 |
| High school—All grades | Embedded American Sign Language | ACC | 137 | 68 | 49.64 |
| High school—All grades | Embedded Speech-to-Text | ACC | 3,352 | 169 | 5.04 |
| High school—All grades | Non-Embedded Print-on-Demand | ACC | 105 | 6 | 5.71 |
| High school—All grades | Embedded Masking | DS | 10,929 | 331 | 3.03 |
| High school—All grades | Embedded Text-to-Speech (English TTS) | DS | 27,482 | 3,575 | 13.01 |

### Practice and Training Tests

Practice and training tests are available publicly to LEA staff, students, parents/guardians, and any other individual for CAST. These tests simulate the experience of the computer-based CAST to allow anyone to experience the assessment. For the 2023–24 school year, accommodated versions of CAST practice and training tests were developed to include all accessibility resources—including braille, closed-captioning, text-to-speech, and audio transcripts—available on the assessment.

Students can access practice and training tests using a web browser. They allow students and administrators to familiarize themselves with the user interface and components of the TDS and help maintain the standardization of test administration. Practice and training tests are available through the Practice and Training Test website linked on the Practice and Training Tests web page on the CAASPP & ELPAC Website.

The practice tests, offered at each grade level or the grade band, were released to prepare students for CAST. These tests more closely simulate CAST’s length and complexity and align with the CAST blueprint.

The grade-level or grade-band specific training tests can be taken by students in all tested grade levels or the grade band. Many unique item types available on the operational assessment are covered in the training tests. The scoring guides for the practice and training tests are available on the Practice and Training Test Resources web page on the CAASPP & ELPAC Website.

*Practice Test Scoring Guides* for each grade level (CDE, 2021d, 2021e, 2021f) and practice test *Constructed Response Annotated Examples* were also available (CDE, 2021a, 2021b, 2021c). The annotated examples provide sample student responses to some of the constructed-response items with accompanying text to provide the rationale for why the response received a score point of 2, 1, or 0.

### Test Security and Confidentiality

For the operational CAST, every person who worked with the assessments, communicated test results, or received testing information was responsible for maintaining the security and confidentiality of the assessments, including CDE staff, ETS staff, ETS subcontractors, LEA staff connected with assessments, school assessment coordinators, students, parents/guardians, and teachers. The ETS Code of Ethics required that all test information, including tangible materials (for example, test items), confidential files (for example, those containing personally identifiable student information), and processes related to test administration (for example, the configurations of secure servers), were kept secure. ETS had systems in place that maintained tight security for test items and test results, as well as for student data. To ensure security for all assessments that ETS develops or handles, ETS maintains an Office of Testing Integrity (OTI), which is described in the next subsection.

All assessments within the CAASPP System, as well as the confidentiality of student information, should be protected to ensure the validity, reliability, and fairness of the results. As stated in *Standard 7.9* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), “The documentation should explain the steps necessary to protect test materials and to prevent inappropriate exchange of information during the test administration session” (p. 128).

This section of the *CAST Technical Report* describes the measures intended to prevent potential test security incidents prior to testing and the actions that were taken to handle security incidents occurring during or after the testing window using the STAIRS process.

#### The ETS Office of Testing Integrity

The OTI is a division of ETS that provides quality-assurance services for all testing programs managed by ETS. This division resides in the ETS legal department. The Office of Professional Standards Compliance at ETS publishes and maintains the *ETS Standards for Quality and Fairness* (2014), which supports OTI goals and activities. The *ETS Standards for Quality and Fairness* provides guidelines to help ETS staff design, develop, and deliver technically sound, fair, and beneficial products and services and help the public and auditors evaluate those products and services.

The OTI mission is to

* prevent test security violations;
* minimize any testing security violations that can impact the fairness of testing;
* minimize and investigate any security breach that threatens the validity of the interpretation of test scores; and
* report on security activities.

The OTI helps prevent misconduct on the part of students and administrators, detects potential misconduct through empirically established indicators, and resolves situations involving misconduct in a fair and equitable way that reflects the laws and professional standards governing the integrity of testing. The OTI also implements policies designed to detect and block technologies used to gain an unfair advantage.

In its pursuit of enforcing secure testing practices, the OTI strives to safeguard the various processes involved in an assessment development and administration cycle. For CAST, those processes included the following:

* Assessment development
* Item and data review
* Item banking
* Transfer of forms and items to the CDE and CAI
* Security of electronic files using a firewall
* Test administration
* Test delivery
* Processing and scoring
* Data management
* Statistical analysis
* Student confidentiality

#### Procedures to Maintain Standardization of Test Security

Test security requires the accounting of all secure materials—including computer-based summative test items and student data—before, during, and after each test administration. The LEA CAASPP coordinator is responsible for keeping all electronic test materials secure, keeping student information confidential, and making sure the site CAASPP coordinators and test administrators are properly trained regarding security policies and procedures.

The site CAASPP coordinator is responsible for mitigating test security incidents at the test site and for reporting incidents to the LEA CAASPP coordinator.

The test administrator is responsible for reporting testing incidents to the site CAASPP coordinator and securely destroying printed and digital media for items and passages generated by the print-on-demand feature of the TDS (CDE, 2024d).

The following measures ensured the security of the CAASPP:

* LEA CAASPP coordinators and site CAASPP coordinators must have electronically signed and submitted a “CAASPP Test Security Agreement for LEA CAASPP coordinators and site CAASPP coordinators” form in TOMS before ETS can grant the coordinators access to TOMS (5 *CCR* Section 859[d]).
* Anyone having access to the testing materials must have electronically signed and submitted a “Test Security Affidavit for Test Examiners, Test Administrators, Proctors, Translators, Scribes, and Any Other Person Having Access to CAASPP Tests” form in TOMS before receiving access to any testing materials (5 *CCR* Section 859[c]).
* Anyone having access to the testing materials but not having access to TOMS must have signed the CAASPP *Test Security Affidavit for Non-TOMS Users*, which was available as a web-based form, before receiving access to any testing materials.

In addition, it was the responsibility of every participant in the CAASPP System to report immediately any violation or suspected violation of test security or confidentiality. The test administrator reported to the site CAASPP coordinator or LEA CAASPP coordinator, who then submitted the incident using the STAIRS/Appeals process. Breach incidents were to be reported by the LEA CAASPP coordinator to the California Technical Assistance Center (CalTAC) and entered into STAIRS within 24 hours of the incident (5 *CCR* Section 859[e]).

#### Test Security Monitoring

The LEA and school testing staff were responsible for maintaining the security and confidentiality of testing materials and devices during the testing window and reporting any irregularities or breaches that occurred. ETS performed site visits and testing procedure audits at randomly selected LEAs and test sites throughout California during the test administration of CAASPP and the ELPAC operational assessments. Audits were performed before, during, and after test administrations to observe adherence to published procedures regarding the handling of testing materials and test administration guidelines.

For Tier 1 audits, 200 randomly selected sites, with their respective LEAs, responded to an online survey. ETS program staff conducted Tier 2 audits of LEAs in virtual meetings on the basis of significant security concerns such as the number of STAIRS cases submitted, communications from the LEA, security concerns reported to the CDE or ETS, or their responses to the audit survey. While there were no Tier 3 audits conducted during the 2023–24 test administration, these would have been conducted in person by a member of the ETS California program team if security issues were deemed to have disregarded rules of testing that resulted in a severe security breach. If, during a Tier 3 audit, the ETS staff member deemed the security issue to be of highest concern, then ETS OTI would be asked to escalate the investigation.

ETS provided a final summary report of audit findings to the CDE at the end of the test administration. ETS program management reported a summary of these findings to the CDE after a site visit.

#### Security of Electronic Files Using a Firewall

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

All electronic applications that are included in TOMS remain protected by the ETS firewall software at all times. Because of the sensitive nature of the student information processed by TOMS, the firewall plays a significant role in maintaining assurance of confidentiality among the users of this information.

Refer to section [*1.9 Systems Overview and Functionality*](#_Systems_Overview_and) in[*Chapter 1: Introduction*](#_Introduction) for more information on TOMS.

#### Transfer of Scores via Secure Data Exchange

Because of the confidential nature of test results, ETS uses secure file transfer protocol (SFTP) and encryption for all data file transfers; test data is never sent via email. SFTP is a method for reliable and exclusive routing of files. Files reside on a password-protected server that only authorized users can access. ETS shares an SFTP server with the CDE. On that site, ETS posts Microsoft Word and Excel files, Adobe Acrobat PDFs, or other document files for the CDE to review; the CDE returns reviewed materials in the same manner. Files are deleted upon retrieval.

The SFTP server is used as a conduit for the transfer of files; secure test data is stored only temporarily on the shared SFTP server. Industry-standard secure protocols are used to transfer test content and student data from the ETS internal data center to any external systems.

For the 2023–24 CAST, ETS entered information about the deliverable into a web form on a SharePoint website when a file was posted. A CDE staff member monitored this log throughout the day for updates to the status of deliverables and downloaded and deleted the file from the SFTP server when its status showed that it had been posted.

#### Data Management in the Secure Database

ETS maintains a secure database to house all student demographic data and assessment results. Information associated with each student has a database relationship to the LEA, school, and grade codes as the data is collected during testing. Only individuals with the appropriate credentials can access the data. ETS builds all interfaces with the most stringent security considerations, including interfaces with data encryption for databases that store test items and student data. ETS applies best and up-to-date security practices, including system-to-system authentication and authorization, in all solution designs.

All stored test content and student data is encrypted. Industry-standard secure protocols are used to transfer test content and student data from the ETS internal data center to any external systems. ETS complies with the Family Educational Rights and Privacy Act (20 *United States Code* [*USC*] § 1232g; 34 *Code of Federal Regulations* Part 99) and the Children’s Online Privacy Protection Act (15 *USC* §§ 6501-6506, P.L. No. 105–277, 112 Stat. 2681–1728).

In TOMS, staff at LEAs and test sites have different levels of access appropriate to the role assigned to them (CDE, 2024e).

#### Statistical Analysis on Secure Servers

During CAASPP testing, ETS information technology staff members retrieve data files from CAI and load those files into a database. The ETS Data Quality Services staff extract the data from the database and perform quality-control procedures (for example, the values of all variables are as expected) before passing files to the ETS statistical analysis group. The statistical analysis staff store the files on secure servers. All staff members involved with the data adhere to the ETS Code of Ethics and the ETS Information Protection Policies to prevent any unauthorized access to data.

#### Student Confidentiality

To meet the requirements of the Every Student Succeeds Act, as well as state requirements, LEAs must collect demographic data about students’ ethnicity, disabilities, parent/guardian education, and so forth during the school year. ETS takes every precaution to prevent any of this information from becoming public or being used for anything other than for testing and score-reporting purposes. These procedures are applied to all documents in which student demographic data appears, such as technical reports.

#### Student Test Results

##### Types of Results

The following deliverables are produced for reporting of CAST:

* Individual student results for computer-based assessments in the California Educator Reporting System (CERS)
* Individual SSRs (electronic)
* Internet reports—available on the CDE Test Results for California’s Assessments website—aggregated by content area and state, county, LEA, or test site

##### Security of Results Files

ETS takes measures to protect files and reports that show students’ scores and reporting levels. ETS is committed to safeguarding all secure information in its possession from unauthorized access, disclosure, modification, or destruction. ETS has strict information security policies in place to protect the confidentiality of both student and client data. 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 for unauthorized access or denial of service.

ETS has many facilities, policies, and procedures to protect computer files. 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 both the ISO 27001 standard for information security and the ISO 22301 standard for business continuity, and conducts disaster recovery exercises annually.

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 ETS Computer Processing Center at all times. Extensive smoke detection and alarm systems, as well as a preaction fire-control system, are installed in the Center.

##### Security of Individual Results

ETS protects individual students’ results during the following conditions:

* Data collection from the TDS
* Scoring
* Transfer of scores by means of secure data exchange
* Reporting
* Posting of aggregate data
* Storage

In addition to protecting the confidentiality of testing materials, The ETS Code of Ethics further prohibits ETS employees from financial misuse, conflicts of interest, and unauthorized appropriation of ETS property and resources. Specific rules are also given to ETS employees and their immediate families who may take an assessment developed by ETS. The ETS OTI verifies that these standards are followed throughout ETS. This verification is conducted, in part, by periodic on-site security audits of departments, with follow-up reports containing recommendations for improvement.

#### Security and Test Administration Incident Reporting System Process

Test security incidents, such as improprieties, irregularities, and breaches, are prohibited behaviors that give a student an unfair advantage or compromise the secure administration of the assessments, which, in turn, compromise the reliability and validity of test results (CDE, 2024b). Whether intentional or unintentional, failure by staff or students to comply with security rules constitutes a test security incident. Test security incidents impact scoring and affect students’ performance on the assessment.

LEA CAASPP coordinators and site CAASPP coordinators ensured that all test security and test administration incidents were documented by following the prompts in TOMS that guided coordinators in their submittal. An Appeal is a request to reset, restore, reopen, invalidate, or grant a grace period extension to a student’s assessment. If an Appeal to a student’s assessment was warranted, TOMS provided additional prompts to file the Appeal.

After a case was submitted, an email containing a case number and next steps was sent to the submitter (and to the LEA CAASPP coordinator, if the case was submitted by the site CAASPP coordinator). The STAIRScase in TOMS provided the LEA CAASPP coordinator, the CDE, and the LEA Outreach Administrator with the opportunity to interact and communicate regarding the STAIRS process (CDE, 2024b).

Prior to the assessment administration, ETS and the CDE agreed that the following types of STAIRS cases would also be forwarded to the CDE:

* Student cheating or accessing unauthorized devices
* Security breach (where a student exposed secure materials)
* Student unable to review previous answers (that is, 20-minute pause rule)

Appeals requests were reviewed by the CDE or an ETS LEA Outreach Administrator. When a request to submit an Appeal was approved, the coordinator received a system-generated email with the Appeal type that was approved (CDE, 2024b).

Types of Appeals available during the 2023–24 CAST administration are described in table 5.3.

Table 5.3 Types of Appeals

|  |  |
| --- | --- |
| **Type of Appeal** | **Description** |
| Reset | Resetting a student’s assessment entailed removing that assessment from the TDS and enabling the student to start a new assessment from the beginning. |
| Invalidate | Invalidated assessments were scored, and scores were provided on the SSR with a note that an irregularity occurred. The student(s) was counted as participating in the calculation of the school’s participation rate for accountability purposes. |
| Reopen | Reopening an assessment allowed a student to access an assessment that had already been submitted or had expired. |
| Restore | Restoring an assessment returned an assessment from the Reset status to its prior status. This action could be performed only on assessments that were reset previously. |
| Grace Period Extension | Permitting a grace period extension allowed the student to review previously answered items upon logging back on to the assessment after expiration of the pause rule. Note that for a performance task, having the test administrator open a new testing session may be all that was needed to continue testing.  A grace period extension was granted only in cases where there was a disruption to a test session, such as a technical difficulty, fire drill, schoolwide power outage, earthquake, or other act beyond the control of the test administrator. |

##### Impropriety

A testing impropriety is an unusual circumstance that has a low impact on the individual or group of students who are testing and has a low risk of potentially affecting student performance on the assessment, test security, or test validity. An example of an impropriety could be if students were making distracting gestures or sounds or talking during the test session that creates a disruption in the test session for other students, or a student left the test room without authorization.

An impropriety can be corrected and contained at a local level. An impropriety should be reported to the LEA CAASPP coordinator and site CAASPP coordinator immediately. The coordinator must report the incident within 24 hours, using the STAIRS/Appeals process in TOMS.

##### Irregularity

A testing irregularity is an unusual circumstance that impacts an individual or a group of students who are testing and may potentially affect student performance on the assessment or impact test security or test validity. An example of an irregularity could be that students were assigned an incorrect designated support or accommodation, or students cheated or provided answers to each other.

These circumstances can be corrected and contained at the local level and submitted using the STAIRS/Appeals process in TOMS. An irregularity must be reported to the LEA CAASPP coordinator and site CAASPP coordinator immediately. The coordinator must report the irregularity within 24 hours, using the online STAIRS/Appeals process in TOMS.

##### Breach

A testing breach is an event that poses a threat to the validity of the assessment. Examples may include such situations as a release of secure materials or a security or system risk. These circumstances have external implications for the CDE and may result in a decision to remove the test item(s) from the available secure item bank.

Breaches require immediate attention; a breach that was due to social media exposure on the part of a student or adult or due to media coverage of an administration was to be escalated to CalTAC via a telephone call from the LEA CAASPP coordinator. Following the call, the site CAASPP coordinator or LEA CAASPP coordinator must report the incident using the online STAIRS/Appeals process in TOMS within 24 hours. All other breaches were to be entered into STAIRS directly.

#### Appeals

For test security incidents reported in STAIRS that resulted in a need to invalidate, restore, or provide a grace period extension for individual computer-based student assessments, the request had to be approved by the CDE. Requests to reset and reopen assessments were processed by an LEA Outreach Administrator.

In most instances, an Appeal was submitted to address a test security breach or irregularity. The LEA CAASPP coordinator or site CAASPP coordinator submitted Appeals in TOMS. All submitted Appeals were available for retrieval and reviewed by LEA and site coordinators within a given organization. An Appeal could be requested only by the LEA CAASPP coordinator or site CAASPP coordinator if prompted while filing a STAIRS case in TOMS (CDE, 2024a). Types of Appeals available during the 2023–24 administration are described in table 5.3.

Table 5.4 presents the number and types of testing issues submitted in STAIRS in the 2023–24 test administration for CAST as well as the number of individual Statewide Student Identifiers (SSIDs) submitted and approved. The most frequently reported incident was for accessibility issues. The number in the *Appeals SSIDs Approved* column is the number of accepted cases that resulted in an Appeal, which may differ from the number in the *Number of Incidents* column because of incorrect entry or other factors.

Table 5.4 Number and Types of Incidents Submitted in STAIRS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Description** | **Appeal Type** | **Number of Incidents** | **Total Number of SSIDs Submitted** | **Appeals SSIDs Approved** |
| Accessibility Issue | Reset | 518 | 1,302 | 1,237 |
| Administered Incorrect Assessment | Reset or Reopen or No Appeal | 103 | 266 | 112 |
| Administration Error | No Appeal | 8 | 0 | 0 |
| Data Entry Issue | Reset or Reopen or Invalidate or No Appeal | 0 | 0 | 0 |
| Expired or Accidentally Submitted Test | Reopen | 178 | 552 | 550 |
| Exposing Secure Materials | Invalidate or No Appeal | 6 | 6 | 6 |
| Incorrect SSID Used | Reset or No Appeal | 31 | 35 | 27 |
| Restore from Reset | Restore | 6 | 7 | 6 |
| Student Cheating or Accessing Unauthorized Devices | Invalidate | 98 | 106 | 99 |
| Student Disruption | No Appeal | 12 | 0 | 0 |
| Technical Issues | Grace Period Extension or No Appeal | 74 | 1,530 | 1,490 |
| Validity Issue | Invalidate or Reset | 12 | 39 | 35 |
| **Totals:** | **N/A** | **1,046** | **3,843** | **3,562** |

Table 5.5 presents the number of Appeals approved and rejected in STAIRS in the 2023–24 administration of CAST, for all grade levels combined.

Table 5.5 Number of Appeals Requested in STAIRS—‍All Grade Levels

|  |  |  |
| --- | --- | --- |
| **Appeal Type** | **Number of Appeals Approved** | **Number of Appeals Rejected** |
| Reset | 1,344 | 95 |
| Invalidate | 140 | 10 |
| Reopen | 577 | 2 |
| Restore | 6 | 1 |
| Grace Period Extension | 1,490 | 40 |
| **Totals:** | **3,557** | **148** |

### References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

*California* *Code of Regulations*,Title 5, Education, Division 1, Chapter 2, Subchapter 3.75, Article 2.

California Department of Education. (2021a). *California Science Test constructed response annotated examples—grade five*. Sacramento, CA: California Department of Education.

California Department of Education. (2021b). *California Science Test constructed response annotated examples—grade eight*. Sacramento, CA: California Department of Education.

California Department of Education. (2021c). *California Science Test constructed response annotated examples—high school*. Sacramento, CA: California Department of Education.

California Department of Education. (2021d). *California Science Test practice test scoring guide for grade eight*. Sacramento, CA: California Department of Education.

California Department of Education. (2021e). *California Science Test practice test scoring guide for grade five*. Sacramento, CA: California Department of Education.

California Department of Education. (2021f). *California Science Test practice test scoring guide for high school*. Sacramento, CA: California Department of Education.

California Department of Education. (2021g). *Science test administration for high school students*. Sacramento, CA: California Department of Education.

California Department of Education. (2023). *California assessment accessibility resources matrix*. Sacramento, CA: California Department of Education.

California Department of Education. (2024a). *CAASPP and ELPAC accessibility guide*. Sacramento, CA: California Department of Education.

California Department of Education. (2024b). *CAASPP and ELPAC security incidents and Appeals procedure guide*.Sacramento, CA: California Department of Education.

California Department of Education. (2024c). *CAASPP and ELPAC technical specifications and configuration guide for online testing*. Sacramento, CA: California Department of Education.

California Department of Education. (2024d). *CAASPP online test administration manual*. Sacramento, CA: California Department of Education.

California Department of Education. (2024e). *Test Operations Management System user guide*. Sacramento, CA: California Department of Education.

Educational Testing Service. (2014). *ETS standards for quality and fairness*. Princeton, NJ: Educational Testing Service.

Every Student Succeeds Act of 2015, 20 U.S.C. 20 § 6301 (2015).

Smarter Balanced Assessment Consortium. (2014). *Smarter Balanced Assessment Consortium: Usability, accessibility, and accommodations implementation guide*.Los Angeles: Smarter Balanced Assessment Consortium.

Smarter Balanced Assessment Consortium. (2018). *Smarter Balanced resources and practices comparison crosswalk*. Los Angeles: Smarter Balanced Assessment Consortium.

Smarter Balanced Assessment Consortium. (2023). *Smarter Balanced Assessment Consortium: Usability, accessibility, and accommodations guidelines*. Los Angeles: Smarter Balanced Assessment Consortium.

Smarter Balanced Assessment Consortium. (2024). *Accessibility strategies* (web page). Los Angeles: Smarter Balanced Assessment Consortium.

### Appendix 5.A: Accessibility Resource Assignment

Table 5.A.1 Accessibility Resource Assignment, Grades Five and Eight—All Tested

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Accessibility Resource** | **Grade 5 Number** | **Grade 5 Pct. of Total** | **Grade 8 Number** | **Grade 8 Pct. of Total** |
| Embedded Accommodation—American Sign Language | 159 | 0.04 | 149 | 0.03 |
| Embedded Accommodation—Braille | 15 | <0.01 | 9 | <0.01 |
| Embedded Accommodation—Speech-to-Text | 16,226 | 3.84 | 8,842 | 2.07 |
| Embedded Accommodation—Word Prediction | 4,366 | 1.03 | 2,851 | 0.67 |
| Non-Embedded Accommodation—Abacus | 325 | 0.08 | 95 | 0.02 |
| Non-Embedded Accommodation—Alternate Response Options | 241 | 0.06 | 135 | 0.03 |
| Non-Embedded Accommodation—Print-on-Demand | 114 | 0.03 | 67 | 0.02 |
| Non-Embedded Accommodation—Speech-to-Text | 11,389 | 2.70 | 7,628 | 1.78 |
| Non-Embedded Accommodation—Word Prediction | 4,696 | 1.11 | 4,154 | 0.97 |
| Embedded Designated Support—Color Contrast | 1,399 | 0.33 | 641 | 0.15 |
| Embedded Designated Support—Masking | 11,933 | 2.83 | 9,967 | 2.33 |
| Embedded Designated Support—Mouse Pointer (Size and Color) | 3,891 | 0.92 | 2,427 | 0.57 |
| Embedded Designated Support—Permissive Mode | 1,511 | 0.36 | 1,202 | 0.28 |
| Embedded Designated Support—Print Size | 2,749 | 0.65 | 1,084 | 0.25 |
| Embedded Designated Support—Stacked Translations | 5,770 | 1.37 | 3,827 | 0.90 |
| Embedded Designated Support—Streamlined Test Interface | 7,281 | 1.72 | 5,931 | 1.39 |
| Embedded Designated Support—Text-to-Speech | 67,034 | 15.87 | 46,824 | 10.95 |
| Embedded Designated Support—Translation Glossary | 5,378 | 1.27 | 4,249 | 0.99 |
| Embedded Designated Support—Turn off Universal Tools | 0 | 0.00 | 0 | 0.00 |
| Non-Embedded Designated Support—100s Number Table | 19,113 | 4.52 | 8,433 | 1.97 |
| Non-Embedded Designated Support—Amplification | 2,520 | 0.60 | 2,108 | 0.49 |
| Non-Embedded Designated Support—Calculator | 10,666 | 2.53 | 19,573 | 4.58 |
| Non-Embedded Designated Support—Color Contrast | 406 | 0.10 | 188 | 0.04 |
| Non-Embedded Designated Support—Color Overlay | 167 | 0.04 | 106 | 0.02 |
| Non-Embedded Designated Support—Magnification | 1,075 | 0.25 | 1,127 | 0.26 |
| Non-Embedded Designated Support—Medical Device | 109 | 0.03 | 105 | 0.02 |
| Non-Embedded Designated Support—Multiplication Table | 27,547 | 6.52 | 18,609 | 4.35 |
| Non-Embedded Designated Support—Noise Buffers | 8,481 | 2.01 | 7,020 | 1.64 |
| Non-Embedded Designated Support—Read Aloud | 12,695 | 3.01 | 7,365 | 1.72 |
| Non-Embedded Designated Support—Read Aloud in Spanish | 0 | 0.00 | 0 | 0.00 |
| Non-Embedded Designated Support—Science Charts | 10,928 | 2.59 | 20,333 | 4.76 |
| Non-Embedded Designated Support—Scribe | 3,650 | 0.86 | 1,402 | 0.33 |
| Non-Embedded Designated Support—Separate Setting | 36,798 | 8.71 | 29,858 | 6.99 |
| Non-Embedded Designated Support—Simplified Test Directions | 32,890 | 7.79 | 26,197 | 6.13 |
| Non-Embedded Designated Support—Translated Test Directions | 1,841 | 0.44 | 1,783 | 0.42 |
| Other—Unlisted Resources | 0 | 0.00 | 0 | 0.00 |
| Other—Designated support or accommodation is in IEP | 44,855 | 10.62 | 42,319 | 9.90 |
| Other—Designated support or accommodation is in Section 504 plan | 2,995 | 0.71 | 2,706 | 0.63 |

Table 5.A.2 Accessibility Resource Assignment, High School Grades—All Tested

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Accessibility Resource** | **Grade 10 Number** | **Grade 10 Pct. of Total** | **Grade 11 Number** | **Grade 11 Pct. of Total** | **Grade 12 Number** | **Grade 12 Pct. of Total** | **High School Number** | **High School Pct. of Total** |
| Embedded Accommodation—American Sign Language | 12 | 0.05 | 92 | 0.03 | 33 | 0.03 | 137 | 0.03 |
| Embedded Accommodation—Braille | 2 | <0.01 | 11 | <0.01 | 1 | <0.01 | 14 | <0.01 |
| Embedded Accommodation—Speech-to-Text | 252 | 1.01 | 2,620 | 0.79 | 480 | 0.48 | 3,352 | 0.73 |
| Embedded Accommodation—Word Prediction | 30 | 0.12 | 791 | 0.24 | 130 | 0.13 | 951 | 0.21 |
| Non-Embedded Accommodation—Abacus | 1 | <0.01 | 43 | 0.01 | 7 | <0.01 | 51 | 0.01 |
| Non-Embedded Accommodation—Alternate Response Options | 2 | <0.01 | 75 | 0.02 | 14 | 0.01 | 91 | 0.02 |
| Non-Embedded Accommodation—Print-on-Demand | 1 | <0.01 | 74 | 0.02 | 30 | 0.03 | 105 | 0.02 |
| Non-Embedded Accommodation—Speech-to-Text | 210 | 0.84 | 3,162 | 0.95 | 570 | 0.57 | 3,942 | 0.86 |
| Non-Embedded Accommodation—Word Prediction | 87 | 0.35 | 1,454 | 0.44 | 288 | 0.29 | 1,829 | 0.40 |
| Embedded Designated Support—Color Contrast | 11 | 0.04 | 2,140 | 0.64 | 118 | 0.12 | 2,269 | 0.50 |
| Embedded Designated Support—Masking | 273 | 1.09 | 8,813 | 2.65 | 1,843 | 1.85 | 10,929 | 2.39 |
| Embedded Designated Support—Mouse Pointer (Size and Color) | 46 | 0.18 | 648 | 0.20 | 1,202 | 1.21 | 1,896 | 0.42 |
| Embedded Designated Support—Permissive Mode | 43 | 0.17 | 297 | 0.09 | 72 | 0.07 | 412 | 0.09 |
| Embedded Designated Support—Print Size | 65 | 0.26 | 638 | 0.19 | 167 | 0.17 | 870 | 0.19 |
| Embedded Designated Support—Stacked Translations | 210 | 0.84 | 2,558 | 0.77 | 641 | 0.64 | 3,409 | 0.75 |
| Embedded Designated Support—Streamlined Test Interface | 372 | 1.49 | 2,454 | 0.74 | 535 | 0.54 | 3,361 | 0.74 |
| Embedded Designated Support—Text-to-Speech | 1,799 | 7.21 | 21,367 | 6.44 | 4,316 | 4.33 | 27,482 | 6.02 |
| Embedded Designated Support—Translation Glossary | 375 | 1.50 | 4,252 | 1.28 | 980 | 0.98 | 5,607 | 1.23 |
| Embedded Designated Support—Turn off Universal Tools | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Non-Embedded Designated Support—100s Number Table | 241 | 0.97 | 4,053 | 1.22 | 420 | 0.42 | 4,714 | 1.03 |
| Non-Embedded Designated Support—Amplification | 11 | 0.04 | 2,283 | 0.69 | 97 | 0.10 | 2,391 | 0.52 |
| Non-Embedded Designated Support—Calculator | 1,233 | 4.94 | 14,306 | 4.31 | 2,915 | 2.93 | 18,454 | 4.04 |
| Non-Embedded Designated Support—Color Contrast | 10 | 0.04 | 145 | 0.04 | 36 | 0.04 | 191 | 0.04 |
| Non-Embedded Designated Support—Color Overlay | 0 | 0.00 | 68 | 0.02 | 9 | <0.01 | 77 | 0.02 |
| Non-Embedded Designated Support—Magnification | 14 | 0.06 | 633 | 0.19 | 108 | 0.11 | 755 | 0.17 |
| Non-Embedded Designated Support—Medical Device | 7 | 0.03 | 47 | 0.01 | 12 | 0.01 | 66 | 0.01 |
| Non-Embedded Designated Support—Multiplication Table | 785 | 3.15 | 7,958 | 2.40 | 1,291 | 1.30 | 10,034 | 2.20 |
| Non-Embedded Designated Support—Noise Buffers | 189 | 0.76 | 3,048 | 0.92 | 622 | 0.62 | 3,859 | 0.85 |
| Non-Embedded Designated Support—Read Aloud | 176 | 0.71 | 3,532 | 1.06 | 695 | 0.70 | 4,403 | 0.96 |
| Non-Embedded Designated Support—Read Aloud in Spanish | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Non-Embedded Designated Support—Science Charts | 2,699 | 10.82 | 15,193 | 4.58 | 3,803 | 3.82 | 21,695 | 4.75 |
| Non-Embedded Designated Support—Scribe | 33 | 0.13 | 479 | 0.14 | 68 | 0.07 | 580 | 0.13 |
| Non-Embedded Designated Support—Separate Setting | 1,677 | 6.73 | 18,579 | 5.60 | 4,390 | 4.41 | 24,646 | 5.40 |
| Non-Embedded Designated Support—Simplified Test Directions | 1,290 | 5.17 | 15,116 | 4.55 | 3,194 | 3.21 | 19,600 | 4.29 |
| Non-Embedded Designated Support—Translated Test Directions | 155 | 0.62 | 1,108 | 0.33 | 407 | 0.41 | 1,670 | 0.37 |
| Other—Unlisted Resources | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Other—Designated support or accommodation is in IEP | 2,364 | 9.48 | 25,107 | 7.56 | 5,927 | 5.95 | 33,398 | 7.32 |
| Other—Designated support or accommodation is in Section 504 plan | 161 | 0.65 | 2,013 | 0.61 | 443 | 0.44 | 2,617 | 0.57 |

## Standard Setting

Standard setting, which also is referred to as achievement level setting, refers to a class of methodologies by which one or more thresholds are used to determine achievement levels. The California Department of Education (CDE) set four achievement levels—*Level 1—Standard Not Met*, *Level 2—Standard Nearly Met*, *Level 3—Standard Met*, and *Level 4—Standard Exceeded*—with three threshold cuts for each grade level and content area.

The CDE and ETS implemented an extensive achievement level–setting process involving software development, item mapping, review panels, committees, workshops, and extensive validity research to set the final thresholds and achievement level descriptors. For detailed information regarding this process, refer to the *Standard Setting Technical Report for the California Science Test* (CDE, 2019).

### Reference

California Department of Education. (2019). *Standard setting technical report for the California Science Test*. Sacramento, CA: California Department of Education.

## Scoring and Reporting

To determine individual students’ scores for the California Science Test (CAST), student item responses were scored, and individual student scores were calculated on the basis of the item responses. In addition, student test scores were aggregated to produce information for schools and local educational agencies (LEAs).

This chapter describes how various types of student responses were scored, as well as the various types of scores and score reports that were produced at the end of administration of CAST.

### CAST Scoring Process

At each tested grade level and the high school grade band, the CAST included selected-response (SR) and constructed-response (CR) items. The SR items are machine-scored, and the CR items are scored by either human scoring or the artificial intelligence (AI) scoring engine.

#### Scoring Constructed-Response Items

All CR items used in the operational (that is, segments A and B) and field test (that is, Segment C) segments of the 2023–24 CAST were scored.

Out of 26 CR items used in segments A and B of the 2023–24 operational forms across the two grade levels and the high school grade band, all of them were AI-scored. The AI models were built and approved for operational use by means of either the data from the 2017–18 field test administration or the following operational administrations, including the 2022–23 test administration. A random sample of approximately 1,800 responses for each operational CR prompt were double-scored by human raters. Double scoring provides a measure of interrater reliability for quality control of the CR item scoring.

Not all students’ responses for the field test CR prompts were scored by human raters. Instead, a random sample of responses drawn from each field test CR item was scored, and those first human ratings were used to support the item analyses, item response theory (IRT) analyses, and construction of AI scoring models for potential use in future operational administrations. A portion of the sampled responses was randomly selected and double-scored (second human ratings) to provide a measure of rating quality.

##### Sampling Process

There are two CR item sampling processes for CAST: one for operational items (for double scoring only) and the other for field test items (for both first and second human ratings). The simple random sampling and, in some cases, stratified sampling methods, are used to ensure representation of demographic student groups.

###### Sampling Process for Double Scoring of Operational Constructed-Response Items

For the 2023–24 test administration, the CAST program double-scored approximately 1,800 student responses for each operational CR item for the purposes of reporting interrater agreement statistics and quality control of CR item scoring. The double scoring also supported the evaluation of AI models.

The sampling for double scoring of operational CR items was conducted randomly, by item, at the time of scoring in the ETS Online Network for Evaluation (ONE). The ETS Psychometric Analysis & Research (PAR) group conducted an evaluation to ensure that the sample represented population demographics, including gender, ethnicity, disability status, English language fluency, and economic status.

###### Sampling Process for Field Test Constructed-Response Items

Each student taking the CAST was expected to receive five operational blocks and one field test block: two operational discrete blocks in Segment A, three operational performance tasks (PTs) in Segment B, and one discrete block or one PT in Segment C (field test segment).

For CR field test prompts in grades five and eight assessments, 1,800 responses per prompt were sampled. For high school, because students may test in grade ten, eleven, or twelve, the percentage of students from each grade level taking the assessment could be drastically different. To account for the uncertainty in sampling composition by grade level, 2,500 responses per prompt were sampled for high school. Out of these sampled responses—for grades five and eight (1,800 for each prompt) and for high school (2,500 for each prompt)—approximately 800 responses per prompt were randomly selected to be double-scored.

The sampling for the field test CR items was conducted for the 2023–24 test administration for all grade levels.

##### Human Scoring Development

CR items need to be human scored to build and evaluate new AI models for field test items and monitor the performance of the approved AI models for operational items. This subsection describes the process by which student responses are scored by human raters.

###### Scoring Rubric

During item development, draft scoring metrics (rubrics) were created with the point scale and descriptions. ETS included these rubrics with the associated items in the internal and external review processes described insection [*3.5 California Educator Review*](#_California_Educator_Review_1). Rubrics were edited as needed on the basis of feedback from the California Department of Education (CDE) and California educators during the item review and range finding processes. Exemplar responses of each score point were provided as benchmarks for scoring guidance.

###### Range Finding

Range finding is the process of identifying student responses that will be used as anchor (benchmark) samples to help ensure that CR items are scored consistently and reliably. For the 2023–24 test administration, range finding was conducted for all grade levels.

Soon after receiving a sufficient volume of CR responses from California schools—approximately 500 responses per CR prompt to be field-tested—ETS began the range finding process by randomly selecting a wide variety of student response samples for each field test prompt. The goal was to ensure sufficient responses at each score point on the rubric to create sets of responses for training, certifying, and monitoring raters during the scoring process. Another part of the range finding process included annotating responses to provide further guidance on why a response received a certain rating.

The following steps describe how the range finding process was implemented:

1. ETS Assessment Development (AD) staff used the rubric (scoring guide) for each item to randomly select and score responses that represent each score point on an item’s rubric. California educators, recruited by the Sacramento County Office of Education (SCOE), applied a subsequent score to initially selected responses. Scored samples needed for various purposes are summarized in table 7.1.

Table 7.1 CAST Sample Selection for Human-Scoring Procedures

|  |  |  |  |
| --- | --- | --- | --- |
| **Sample Type** | **Purpose** | **Number of Sets and Samples in Sets** | **Configuration of Sets** |
| Certification | Certification samples for verifying scoring accuracy of potential raters and scoring leaders | Two sets of 10 samples for one high school 2-point prompt | Two to four samples for each score point represented per set |
| Training | Training samples with annotations for rater training and scoring practice | One set of 10 samples per grade level for each prompt | Two to four samples for each score point per set |
| Benchmarks | Benchmark samples with annotations that represent exemplar responses at each score point on the rubric | One set of eight samples per unique prompt per grade level | Two to three samples for each score point |
| Calibration | Calibration samples for evaluating rater scoring performance on specific prompts | Two sets of five samples per set for each prompt | One to three samples for each score point per set |
| Validity | Validity samples inserted into rater’s scoring queue to monitor the quality of scoring | One set of 20 samples per prompt | Four to 10 samples for each score point |

1. California educators were invited by SCOE to apply to participate in the 2023–24 Range Finding Meeting. From the applicants, SCOE compiled a list of proposed participants diverse in race, ethnicity, and gender and representing a variety of types of schools (rural, suburban, urban; large and small) and regions of the state. The list was sent to the CDE for approval. Twenty-six California educators were approved to participate—8 for grade five, 12 for grade eight, and 6 for high school—with a minimum of 60 percent of educators having science teaching experience.
2. ETS AD staff trained the educators on how to score the sample responses using the ONE scoring system. The educators for each grade level or the high school grade band (that is, grade ten, eleven, or twelve) were assigned to one group. Each group was assigned a set of field test prompts for that grade level or grade band, and the educators in each group scored the samples for each of their assigned prompts. The educators also completed a feedback survey after scoring the samples for each prompt. As a result, each sample response had a minimum of three educator scores.
3. The AD science lead for each grade level or the high school grade band reviewed the educator scores and feedback. AD staff compiled the written and verbal feedback from the educators and provided it to the CDE prior to follow-up virtual meetings with the educators. The purpose of the follow-up meetings was to discuss samples that receive discrepant scores from the educators. The AD science leads identified samples with discrepant scores and facilitated a meeting with each group of educators to reconcile the discrepant scores. CDE staff observed these virtual meetings.
4. On the basis of the educator scores and feedback, AD staff selected responses to include in the required sets for each field test prompt (that is, certification, benchmark, training, calibration, and validity, as summarized in table 7.1). Samples in the benchmark and training sets were annotated. Annotations are short explanations as to why a response earns a particular rating. Annotations help raters make explicit connections between the scoring guide (rubric) and the responses and facilitate the careful and accurate scoring of responses. CDE staff reviewed the proposed benchmark sets for each prompt in the ONE system and approved the sets in reconciliation meetings with ETS. AD staff then created all final sets in the ONE system and used these sets as part of a system of training and controls for verifying the quality and consistency of scoring.

###### Rater Recruitment and Certification Process

Several weeks prior to the start of CR scoring, ETS recruited a pool of eligible CAST raters from invited California science teachers as well as from the current California Assessment of Student Performance and Progress (CAASPP) Smarter Balanced pool of eligible raters from California. All CAST raters were required to have a bachelor’s degree to be eligible to attempt certification. The scoring pool included California educators as well as other raters representing a variety of backgrounds in business, education, and other fields.

Table 7.2 shows the characteristics of the CAST raters. Among the 246 raters scoring for the 2023–24 CAST, 210 responded to the survey, 142 had teaching experience in science, 23 were fluent in Spanish, 166 had experience teaching in a kindergarten through grade twelve (K–12) school, and 25 currently worked in a K–12 school in California. Note that numbers may not match the totals because participants may meet multiple characteristics (such as a current California teacher who had experience teaching science and was fluent in Spanish). The information is self-reported and may not reflect all of their experience.

Table 7.2 Summary of Characteristics of Human Raters Scoring CAST

|  |  |
| --- | --- |
| **Characteristic** | **N** |
| Experience teaching in science | 142 |
| Fluent in Spanish | 23 |
| Experience teaching in a K–12 school | 166 |
| Currently works in a K–12 school in California | 25 |
| Others—Not meeting any of the previous criteria | 27 |
| No response to survey | 36 |

Certification served as an initial screening to ensure that the ETS Scoring and Reporting Operations (SRO) team had a sufficient number of qualified raters in place to meet the demands of scoring. One 2-point prompt (that is, a response that can earn 2, 1, or 0 points) selected from among the operational prompts was used for certification. Training samples were provided for the rater to review and practice rating before attempting certification. If a rater passed certification on the prompt, the rater was eligible to calibrate on the grade-level or grade-band-specific prompts once scoring began.

Raters were required to achieve an 80 percent exact match to the CDE-approved rating for the responses on at least one of the certification sets to be eligible for calibration on a specific test prompt. If raters did not pass either certification set, they were excused from scoring the 2023–24 CAST items.

###### Rater and Scoring Leader Training

ETS selected scoring leaders to oversee a group of raters during the scoring process. Scoring leaders were experienced raters who had demonstrated high scoring accuracy from previous scoring projects at ETS and were invited to act as a scoring leader on a project. The scoring leader backread (read behind), guided, and retrained raters as needed. Each scoring leader monitored a small group of raters on a shift, usually up to 10 raters, to assist SRO with scoring quality.

###### Training for Scoring Leaders

Scoring leaders completed two online training modules using the Learning Management System. The purpose of the training was to discuss the duties of scoring leaders, how to monitor raters using the ONE scoring system, and to provide rater feedback. Scoring leaders were also directed to the training material available in ONE to familiarize themselves with prompt-specific guidance.

###### Training for Raters

Training for raters occurred within the ONE system. Raters were provided with ONE system training documents as well as program-specific information that they could refer to at any time. Prior to attempting calibration, raters were given a window of time to review all training materials in the system and practice scoring using the prescored training sets. After raters completed a training set, they were provided with annotations for each response as a rationale for the rating assigned.

The scoring training provided for each potential rater was designed using CDE-approved materials developed by ETS and followed the three-step progression described as follows:

Step One: Review the Scoring Guide and Benchmarks

Training for scoring began with an overview of the scoring guide, or rubric, and benchmarks. In the ONE system, the rubric was accessed through a tab called [**Scoring Guide**]. The benchmarks, also called anchors, were accessed in ONE through the [**Benchmarks**] tab. The benchmarks had annotations associated with them to call the rater’s attention to specific content in the sample responses.

Step Two: Score Training Sets

After orientation to the scoring guide and the benchmark functions, raters progressed through an online content training in the ONE system, in which they reviewed several sets of sample responses, assigned scores, and received feedback on their scores on the basis of the CDE-approved rating for each response and applicable supporting annotation. Training sets, also called feedback sets, were samples of responses that provided the rater annotations after each sample was scored. The feedback sets for the 2023–24 CAST administration contained a mixed set of sample responses for each score point on the rubric. When raters completed the feedback sets, they could attempt calibration.

Step Three: Set Calibration

Calibration is a system-supported control to ensure raters meet a specified standard of accuracy when scoring a series of prescored responses. Raters calibrated before they were allowed to score, meaning they scored a certain percentage of responses accurately from a set of responses called a calibration set. The passing percentage was determined by the program and was based on score scale (the number of possible scores that could be given) and the number of responses in a set. For CAST, the passing criteria is set at 80 percent exact and zero discrepant.

In general, calibration can be put in place at the beginning of a four- or eight-hour scoring shift prior to starting a new grade level or new prompt or at specified intervals during a scoring window. Raters typically are allowed two chances to calibrate successfully. If raters meet the standard on the first attempt, they proceed directly to scoring responses. If raters are unsuccessful, they may review training sets and attempt to calibrate again with a new calibration set. If they are unsuccessful after both attempts, they are dismissed from that scoring shift.

Calibration can be used as a means to control rater and group drift, which are changes in behavior that affect scoring accuracy between test administrations. Calibration can be used throughout a scoring season (for example, January through July) to check scoring accuracy on a prescored set of responses. For the 2023–24 CAST, raters were calibrated once for each CR item during a three-day scoring period.

For the 2023–24 CAST administration, raters were permitted to score any prompt for a grade level if they passed calibration on their first prompt with a 90 percent exact match for items that are scored 0 or 1 point or an 80 percent match for items that are scored 0, 1, or 2 points.

###### Scoring Rules and Processes

ETS implemented the following scoring rules and processes for CAST operational and field test scoring:

* Operational responses were scored via both human and AI scoring.
* Approximately 1,800 responses per item were double-scored as part of continuous quality management. Raters were not aware when a second scoring was occurring and did not have access to the first score.
* Field test responses were scored only via human scoring.
* Approximately 800 responses for grade five, grade eight, and high school were double-scored to facilitate the building of AI scoring models. Raters were not aware when a second scoring was occurring and did not have access to the first score.
* For field test items only, ETS psychometric staff provided a sampling plan that included the responses selected to be scored. Refer to subsection [*7.1.1.1.2 Sampling Process for Field Test Constructed-Response Items*](#_Sampling_Process_for) for the sampling plan. The sampling plan was uploaded to ONE to activate the responses for scoring.

###### Scoring Monitoring and Quality Management

In addition to the calibration function described previously, raters were monitored closely for the quality of their scoring throughout the scoring window. During a scoring shift, scoring leaders read behind raters at a rate of 10 percent or more of the responses scored by each individual rater to determine whether raters were applying the scoring guide and benchmarks accurately and consistently. When necessary, the scoring leader redirected the rater by referencing the rubric, benchmarks, or both the rubric and benchmarks to explain why a response should have received a different score. When a rater was scoring inconsistently, the backreading proportion might have been more than 10 percent.

Prescored responses from validity sets were also inserted into the rater’s queue for every 10 responses scored. These were inserted in random positions and not fixed, so a rater was unaware which response was a validity response. The ETS CR Performance Measures and Analytics group, in conjunction with AD, reviewed the statistics on the validity responses daily to determine whether raters needed retraining.

The ONE system offers a comprehensive set of tools that the scoring leaders and scoring management staff used to monitor the progress and accuracy of individual raters and raters in the aggregate. ONE generates reports on rater productivity and performance that show the number of responses a rater scored during a shift and how two raters scored the same response (that is, interrater reliability).

###### Interrater Reliability for Operational Items

The ONE system captured interrater reliability by monitoring data for responses that were double-scored. Approximately 1,800 CAST responses per item were double-scored. The statistics included the percentage agreement between the two raters, kappa, quadratic-weighted kappa (QWK), and standardized mean difference (SMD). For detailed descriptions of these statistics, refer to subsection [*8.6.7 Interrater Agreement*](#_Interrater_Agreement) in [*Chapter 8: Psychometric Analyses*](#_Psychometric_Analyses). Scoring management reviewed the interrater reliability statistics for each prompt to determine whether there were any issues that needed to be addressed during scoring.

CAST used the following flagging criteria when identifying operational items to be reviewed for potential elimination after scoring was completed. ETS monitored CAST activity throughout the scoring period and adjusted the training and scoring processes. ETS continually monitors these processes and makes improvements as needed.

Polytomous operational items that were AI-scored were flagged if any of the following conditions occurred when AI scores were compared to human ratings:

* QWK < 0.70
* Absolute value of the SMD > 0.15

Dichotomous items were flagged if either of the following conditions occurred:

* Exact agreement < 0.80
* QWK < 0.70

The interrater reliability statistics are shown in [appendix 8.J](#_Appendix_8.J:_Interrater), table 8.J.1 through table 8.J.3, for the operational CR items that were double-scored by AI and human raters. These tables show that the QWK ranged from 0.68 to 0.94 for all items. Out of 26 items, 25 items have QWKs that are greater than or equal to 0.7, which is the desired threshold for interrater agreement. Only one item has interrater agreement slightly lower than the desired threshold, with a QWK value of 0.68. The absolute values of the SMD are smaller than 0.15 for all 26 items, ranging from 0.00 to 0.09.

Table 7.3 shows the number of items flagged by grade level and the high school grade band. All operational items used in the 2023–24 forms are two-point items. Therefore, “N/A” is indicated in the *Flagged Dichotomous Items* column. There was one flagged item among the 26 operational items across the two grade levels and the high school grade band. Flagged items were subsequently reviewed by content specialists.

Table 7.3 Number of Operational CR Items Flagged

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grade Level or Grade Band | Flagged Polytomous Items | Flagged Dichotomous Items | Total Flagged Items | Total Number of Scored Items | Percent Flagged |
| Grade 5 | 0 | N/A | 0 | 8 | 0 |
| Grade 8 | 0 | N/A | 0 | 9 | 0 |
| High school | 1 | N/A | 1 | 9 | 11 |

The evaluation of CR items with new, approved, AI models are presented in subsection [*7.1.1.3.3 Model Evaluation*](#_Model_Evaluation).

###### Validity Responses and Sets

High interrater reliability is an important goal, and the analysis of related data helps to identify errant scoring. However, validity responses and sets are the most important tools in ensuring scoring accuracy.

Unlike interrater data, which shows a comparison of one rater versus another, validity data indicates the rater’s ongoing ability to match CDE-approved scores when scoring prescored validity responses that are indistinguishable from live responses.

ETS used sample responses approved during the range finding process to create an initial set of 20 validity responses per prompt to represent all points across the score scale. ETS estimated 20 validity responses per prompt would be sufficient for the scoring window.

Review of incorrectly scored validity responses was an ongoing process that alerted scoring leaders to specific needs for monitoring and retraining. Routine procedures included focused backreading that could lead to one-on-one retraining sessions between scoring leaders and individual raters. Additionally, scoring leaders and ETS AD staff worked together to identify any trends in errant scoring patterns to determine whether a broader retraining effort would be beneficial, such as the creation of an additional training set to reanchor, or refocus, the group in the accurate application of a particular aspect of the scoring guide.

ETS AD and CR Scoring Systems and Capabilities staff reviewed raters’ scoring patterns and made judgment calls on whether to dismiss a rater. Raters who were unable to maintain an adequate standard of accuracy after retraining were disqualified from scoring the item. When a rater was dismissed, ETS scoring leadership reviewed the rater’s scoring patterns to determine whether all scores assigned by the rater during the time period in question should be nullified and the responses routed for rescoring.

Features such as backreading, interrater reliability reporting functions, and validity response insertion and reporting functions allowed scoring leaders to quickly identify inaccurate scoring patterns and take appropriate corrective actions.

##### Artificial Intelligence Model Building

In 2024, ETS built models for 37 field test items (14 for grade five, 16 for grade eight, and 7 for high school) that were human-scored during the 2023–24 test administration. Of the 37 AI models that were built, 35 were approved for operational scoring. The evaluation process of the AI models is presented in subsection [*7.1.1.3.3 Model Evaluation*](#_Model_Evaluation).

###### Data Collection

After the CAST administration, ETS collected a sample of students’ responses to 37 CR items with human score(s) assigned, as described in subsection [*7.1.1.1 Sampling Process*](#_Sampling_Process).

###### Model Training

At ETS, the steps to build AI scoring models for scoring text-based responses involved the automatic extraction and modeling of linguistic features. Natural language processing techniques were used to extract construct-relevant linguistic features from a set of human-scored responses. Using the linguistic features extracted from the data, statistical models were built to predict the scores that human raters would assign to that response. Statistical modeling methods included, for example, multiple linear regression and support vector machines (SVMs).[[8]](#footnote-9) Each model was built using a tenfold cross-validation method that randomly split the entire dataset for an item into 10 subsets. Nine instances of the data are used to train the model, while the tenth instance is used to test the predictive ability of the model. The subsets are rotated so the final model for each item uses the entire dataset for training and testing.

Each model then went through an evaluation stage with multiple statistical criteria, such as Pearson’s *r* and QWK, using the predictions from each testing instance. The evaluations performed are reported in the next subsection.

Figure 7.1 depicts a cycle chart illustrating the key stages and steps in the AI model-building and evaluation process. At a high level, the three-square diagram indicates that the scoring model is the outcome of model building and serves as input for model evaluation. In the content squares, the three labeled scored responses in the funnel represent sample responses rated by humans with scores of 1 for response 1, 1 for response 2, and 2 for response 3. The two rounded rectangles at the bottom represent the two steps of the model-building process, which are extract linguistic features and statistical modeling. An arrow indicates the direction of these human-scored responses to natural language processing tools for extracting linguistic features as the first step. Another arrow directs the process to the next step, statistical modeling. Here, the model-building process ends. The resulting scoring model from the previous steps—the second square, which is indicated as an arrow from the model-building steps—then proceeds to model evaluation (the third square).

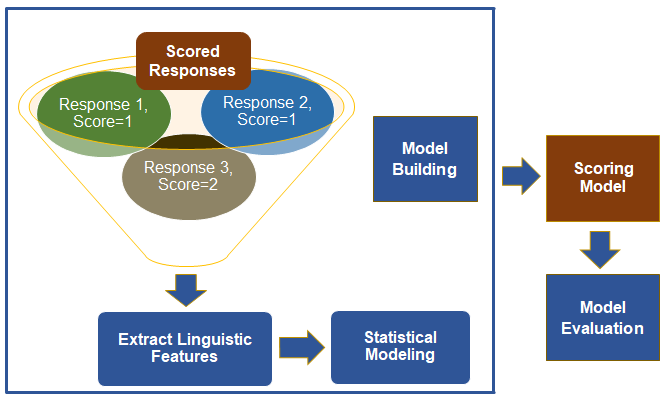


Figure 7.1 AI model-building and evaluation process

###### Model Evaluation

One of the important factors in building AI scoring models with good performance is the use of data with reliable human scores. A commonly used indicator for evaluating human scoring reliability is to use multiple raters on a sufficiently large sample size (at least 800 responses) and evaluate the extent to which they agree with each other.

Each item response had either one or two human ratings: first and second human ratings. The second human ratings were available only on those randomly selected item responses that were double-scored. The first human ratings were used to build and evaluate an AI model, and the second human ratings were used to validate the first human ratings. The evaluation of an AI model includes human–human agreement (exact agreement, adjacent agreement, human–human QWK); human–AI agreement (human–AI QWK); and the comparison of the two types of agreement (degradation of QWK): SMD and proportional reduction in mean squared error (PRMSE). High human–human agreement indicates that the human ratings used to build the AI model are reliable. High human–AI agreement that is similar to the human–human agreement for the item indicates that the AI model performs at a level comparable to humans.

ETS updated the AI criterion to recommend field test items for AI scoring in 2023–24 (refer to subsection [*11.6.1 Updated Artificial Intelligence Scoring Criteria*](#_Updated_Artificial_Intelligence) for details). Thresholds for these measures of AI scoring quality are given in table 7.4 and are commonly applied to studies of AI scoring engines.

Table 7.4 Performance Criteria for Acceptance for AI Scoring

|  |  |
| --- | --- |
| **Statistical Metrics** | **Performance Criteria** |
| QWK (H:H) | QWKH:H ≥ 0.60 |
| QWK (H:AI) | QWKH:AI ≥ 0.70 |
| Degradation of QWK (H:H−H:M) | QWKDeg ≤ 0.10 |
| PRMSE | PRMSE ≥ 0.70 |
| Overall SMD | |SMD| ≤ 0.15 |

In addition to the overall SMD, student group SMDs between the AI ratings and the human ratings are also evaluated. However, student group |SMD| < 0.10 is to be used as a threshold and not as an acceptance criterion. Any items with SMD equal to or greater than 0.10 would be flagged for further investigation.

#### Scoring for Selected-Response Items

The 2023–24 CAST included machine-scorable, traditional multiple-choice items, and technology-enhanced items that were scored by the Test Delivery System (TDS). In the TDS, responses to the test forms were compared with the answer keys or scoring rubrics embedded in the TDS to determine the score points. A real-time, quality-monitoring component was built into the TDS. After an assessment was administered to a student, the TDS passed the resulting data to the Quality Assurance System to ensure a score from the machine-scoring system was accurate. The details of quality control are provided in section [*9.5* *Quality Control of Scoring*](#_Quality_Control_of_2).

### Student Test Scores

ETS developed two parallel scoring systems to produce students’ scores: the Enterprise Score Key Management (eSKM) scoring system, which scores and delivers individual students’ scores to the ETS reporting system; and a parallel scoring program used to validate student score calculations. The ETS PAR team used this program to compute individual students’ scores. The two scoring systems independently applied the same scoring algorithms and specifications. ETS psychometricians verified the eSKM scoring by comparing all individual student scores from PAR and resolving any discrepancies. This internal quality-control step is in place to verify the accuracy of scoring. Students’ scores were reported only when the two parallel systems produced results with acceptable tolerance. Because of the varied optimization algorithms used among systems, a slight difference in the theta estimate is to be expected. An absolute difference in the theta estimates smaller than 0.001 is considered acceptable. A scale score of 1-point difference caused by the small theta difference is also considered acceptable.

All scores must comply with the ETS scoring specifications and the parallel scoring process to ensure the quality and accuracy of scoring and to support the transfer of scores into the database of the student records scoring system, the Test Operations Management System (TOMS).

#### Theta Scores

The 2023–24 CAST scale scores were preequated. The student’s theta score was computed by the inverse test characteristic curve method (Stocking, 1996) via an iterative process. Refer to section [*8.4 Item Response Theory Analyses*](#_Item_Response_Theory) for more details on the IRT models and calibration. This method transformed the sum of the student’s item scores into an ability estimate. That estimate is the ability level at which the sum of the expected scores on the items that the student took is equal to the sum of the raw scores that the student actually earned on those items. The range of theta scores is −4 to 4.

The same method was used to estimate a student’s theta score for both the overall assessment and domain scores.

The individual student theta score distributions are presented in table 7.A.1 through table 7.A.6 in [appendix 7.A](#_Appendix_7.A:_Overall) for grades five, eight, ten, eleven, and twelve, as well as for the high school grade band as a whole.

#### Scale Scores for the Total Assessment

CAST uses an IRT model to estimate students’ abilities (that is, theta scores) and then uses the IRT true score equating method to convert the theta scores to number-right (NR) scores on a base form, which is composed of 100 Rasch items with a difficulty of 0. A total theta score is converted to an NR score using equation 7.1. *Refer to the* [*Alternative Text for Equation 7.1*](#_Alternative_Text_for) *for a description of this equation.*

 (7.1)

where is the NR score for student *j*, and

 is the estimated theta score for student *j*.

Because all forms are equated to one base form, the NR scores account for the form difficulty differences and are comparable across forms. Because the NR scores can easily be misinterpreted as raw scores, a transformation is needed to convert them to scale scores to facilitate score interpretation. Table 7.5 shows the scaling constants for the linear transformation of an NR score to a scale score.

Table 7.5 Scaling Constants

|  |  |  |
| --- | --- | --- |
| **Grade Level or Grade Band** | **Slope** | **Intercept** |
| Grade 5 | 1.0081 | 151 |
| Grade 8 | 1.0081 | 351 |
| High school | 1.0081 | 551 |

The transformation constants are derived by mapping the lowest obtainable NR score to the lowest obtainable scale score (LOSS) plus one, and the highest obtainable NR score to the highest obtainable scale score. The solutions to these linear equations are the transformation constants as shown in table 7.5.

The ranges of the reporting scale scores are 150–250, 350–450, and 550–650 for grade five, grade eight, and high school, respectively.

CAST reports scale scores for the total assessment for students who have answered at least 10 items. Those students who did not answer any items on the assessment received NS (no score) in their score report. Those who answered one to nine items received the LOSS. The LOSS for grade five, grade eight, and high school is 150, 350, and 550, respectively.

The CAST is only considered “complete” if a student responds to a minimum number of operational items for the total assessment. The minimum number of operational items required to fulfill the CAST completion requirement for the 2023–24 administration was 43, 44, and 49 for grade five, grade eight, and high school, respectively.

For students who answered at least 10 items but did not complete the assessment (that is, answered fewer items than the requirements), a proportional adjustment on NR is used to provide an equitable score to all students. The amount of adjustment for incomplete test takers is proportional to the fraction of the assessment completed.

Table 7.6 lists the required minimum number of items for students to receive scores for each domain. The domain scores will not be reported for students who failed to meet the requirements.

Table 7.6 Required Number of Items for Reporting Domain Scores

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** |
| Grade 5 | 10 | 10 | 10 |
| Grade 8 | 9 | 10 | 10 |
| High school | 10 | 10 | 12 |

### Summary Statistics

To provide meaningful results to interest holders, test scores for a given grade level and content area are aggregated at the school, LEA or direct funded charter school, county, and state levels. The aggregated scores are generated both for selected groups and for the population. Subsection [*7.2.2* *Scale Scores for the Total Assessment*](#_Scale_Scores_for) presents aggregated results at the state level for computer-based CAST scores. These results are presented in table 7.7, table 7.9, figure 7.2, and table 7.11 through table 7.13. Score aggregation includes only students with valid scores; refer to subsection [*7.4.2 Special Cases*](#_Special_Cases_1)for more information.

Table 7.7 shows the mean and standard deviation (SD) of both scale scores and theta scores for CAST.

Table 7.7 Mean and SD of Theta Scores and Scale Scores

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Number of Students** | **Scale Score Mean** | **Scale Score SD** | **Theta Score Mean** | **Theta Score SD** |
| Grade 5 | 422,398 | 201 | 22.2 | 0.03 | 1.08 |
| Grade 8 | 427,448 | 401 | 22.2 | 0.01 | 1.09 |
| High school—Grade 10 | 24,935 | 599 | 20.3 | −0.09 | 0.96 |
| High school—Grade 11 | 331,949 | 602 | 21.8 | 0.08 | 1.05 |
| High school—Grade 12 | 99,602 | 599 | 21.9 | −0.08 | 1.06 |
| High school—All grades | 456,486 | 601 | 21.8 | 0.03 | 1.05 |

Individual student scale score distributions are presented in table 7.B.1 through table 7.B.6 in [appendix 7.B](#_Appendix_7.B:_Scale) for grades five, eight, ten, eleven, and twelve, as well as for the high school grade band.

#### Total-Test Achievement Levels

A standard setting was conducted after the 2018–19 first operational test administration and achievement levels were established. Student performance on the reporting scale is designated into one of four achievement levels:

* **Level 1—Standard Not Met:** Student demonstrates a minimal understanding of and ability to apply the knowledge and skills associated with the performance expectations (PEs) of the California Next Generation Science Standards (CA NGSS).
* **Level 2—Standard Nearly Met:** Student demonstrates a partial understanding of and ability to apply the knowledge and skills associated with the PEs of the CA NGSS.
* **Level 3—Standard Met:** Student demonstrates an adequate understanding of and ability to apply the knowledge and skills associated with the PEs of the CA NGSS.
* **Level 4—Standard Exceeded:** Student demonstrates a thorough understanding of and ability to apply the knowledge and skills associated with the PEs of the CA NGSS.

The scale score ranges for achievement levels are shown in table 7.8.

Table 7.8 Scale Score Ranges for Achievement Levels

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Standard Not Met** | **Standard Nearly Met** | **Standard Met** | **Standard Exceeded** |
| Grade 5 | 150–178 | 179–213 | 214–230 | 231–250 |
| Grade 8 | 350–377 | 378–414 | 415–432 | 433–450 |
| High school | 550–575 | 576–614 | 615–635 | 636–650 |

Percentages of students in each achievement level are in table 7.9, and their graphic representation is displayed in figure 7.2, which immediately follows.

Table 7.9 Percent of Students in Each Achievement Level for Total Scores

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Number of Students Tested** | **Percent in Achievement Level Standard Not Met** | **Percent in Achievement Level Standard Nearly Met** | **Percent in Achievement Level Standard Met** | **Percent in Achievement Level Exceeded** | **Percent in Achievement Level Met or Exceeded** |
| Grade 5 | 422,398 | 18.48 | 48.84 | 20.50 | 12.18 | 32.68 |
| Grade 8 | 427,448 | 16.46 | 54.59 | 18.94 | 10.01 | 28.95 |
| High school—Grade 10 | 24,935 | 11.73 | 63.64 | 19.71 | 4.92 | 24.63 |
| High school—Grade 11 | 331,949 | 10.91 | 57.00 | 24.04 | 8.04 | 32.08 |
| High school—Grade 12 | 99,602 | 14.35 | 58.92 | 19.71 | 7.02 | 26.73 |
| High school—All grades | 456,486 | 11.71 | 57.79 | 22.86 | 7.65 | 30.51 |

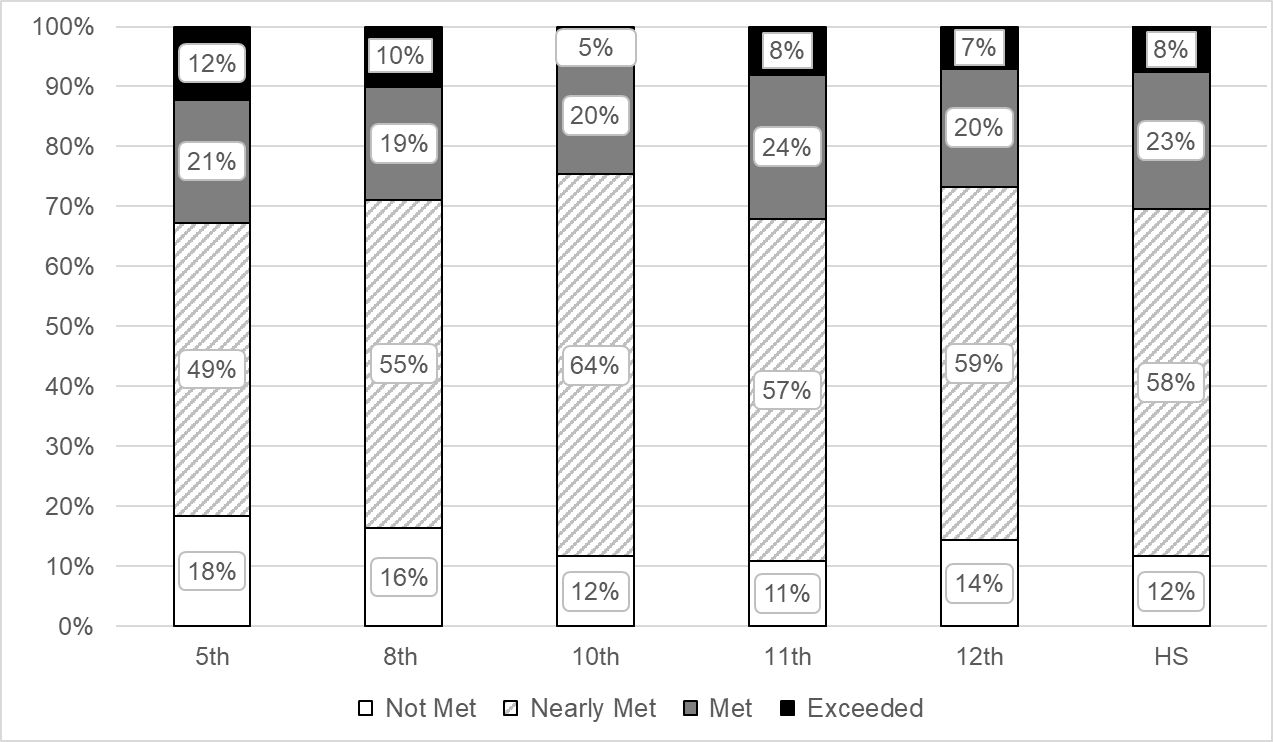


Figure 7.2 Percentage of achievement levels

Overall, 32.68, 28.95, and 30.51 percent of students in grades five and eight and in the high school grade band met or exceeded the standards, respectively. For high school students, the percentage was the lowest for grade ten and the highest for grade eleven.

#### Domain Performance Levels

Science domain performance levels are reported for students who have met the minimum number of item requirements for the science domain as listed in table 7.6. Students might receive science domain performance levels for some domain(s) but not the others depending on the number of items they completed for different domains. Science domain performance levels are not reported for students who received the LOSS (that is, answered fewer than 10 items for the total assessment).

There are three domains for each assessment: Earth and Space Sciences, Life Sciences, and Physical Sciences. (Items aligned with PEs from the Engineering, Technology, and Application of Science subdomain are assigned a science content domain based on the context of the item.) Detailed descriptions of the three performance levels for each science domain are shown in table 7.10.

Table 7.10 Description of Science Domain Performance Levels

|  |  |  |  |
| --- | --- | --- | --- |
| **Science Domain** | **Below Standard** | **Near Standard** | **Above Standard** |
| **Earth and Space Sciences** | The student demonstrates minimal understanding of and ability to apply the knowledge and skills associated with the core ideas, concepts, and practices in Earth and Spaces Sciences, which focus on Earth’s place in the universe, Earth’s systems, and Earth and human activity. | The student demonstrates some understanding of and ability to apply the knowledge and skills associated with the core ideas, concepts, and practices in the Earth and Space Sciences, which focus on Earth’s place in the universe, Earth’s systems, and Earth and human activity. | The student demonstrates a thorough understanding of and ability to apply the knowledge and skills associated with the core ideas, concepts, and practices in the Earth and Spaces Sciences, which focus on Earth’s place in the universe, Earth’s systems, and Earth and human activity. |
| **Life Sciences** | The student demonstrates minimal understanding of and ability to apply the knowledge and skills associated with the core ideas, concepts, and practices in the Life Sciences, which focus on structures and processes in living things, ecosystems, heredity, and biological evolution. | The student demonstrates some understanding of and ability to apply the knowledge and skills associated with the core ideas, concepts, and practices in the Life Sciences, which focus on structures and processes in living things, ecosystems, heredity, and biological evolution. | The student demonstrates a thorough understanding of and ability to apply the knowledge and skills associated with the core ideas, concepts, and practices in the Life Sciences, which focus on structures and processes in living things, ecosystems, heredity, and biological evolution. |
| **Physical Sciences** | The student demonstrates minimal understanding of and ability to apply the knowledge and skills associated with the core ideas, concepts, and practices in Physical Sciences, which focus on matter and its interactions, motion and stability, energy, and waves and their applications. | The student demonstrates some understanding of and ability to apply the knowledge and skills associated with the core ideas, concepts, and practices in the Physical Sciences, which focus on matter and its interactions, motion and stability, energy, and waves and their applications. | The student demonstrates a thorough understanding of and ability to apply the knowledge and skills associated with the core ideas, concepts, and practices in the Physical Sciences, which focus on matter and its interactions, motion and stability, energy, and waves and their applications. |

A student is assigned to one of the three performance levels for a science domain according to the following rules.

Place in the Below Standard level is presented in equation 7.2. *Refer to the* [*Alternative Text for Equation 7.2*](#_Alternative_Text_for_5) *for a description of this equation.*

Equation 7.2; a link to the long description for this equation is found in the preceding paragraph. (7.2)

Place in the Near Standard level is presented in equation 7.3. *Refer to the* [*Alternative Text for Equation 7.3*](#_Alternative_Text_for_6) *for a description of this equation.*

 (7.3)

Place in the Above Standard level is presented in equation 7.4. *Refer to the* [*Alternative Text for Equation 7.4*](#_Alternative_Text_for_14) *for a description of this equation.*

 (7.4)

where,

*θd* is a science domain theta score,

*θL3* is the level 3 theta threshold score of the total assessment, and

 is the conditional standard error of measurement (CSEM) for students with science domain score *θd* and is calculated using equation 8.16.

The domain scale cut scores were established using this method in the 2018–19 test administration. The domain scale cut scores were then reevaluated using the 2021–22 test administration data because of the revised test blueprint and the significant increase in item information. The adjusted domain cut scores better reflected the revised test blueprints. The updated cut scores also reflected the reduced errors in measurement given increased item quality (item discrimination) over time. The CDE approved this change, and the updated domain cut scores were implemented in 2021–22.

Table 7.11 through table 7.13 provide the adjusted domain scale cut scores to align with the revised blueprint.

Table 7.11 Scale Score Ranges in Each Performance Level for the Earth and Space Sciences Domain

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Below Standard** | **Near Standard** | **Above Standard** |
| Grade 5 | 150–198 | 199–233 | 234–250 |
| Grade 8 | 350–393 | 394–435 | 436–450 |
| High school | 550–598 | 599–631 | 632–650 |

Table 7.12 Scale Score Ranges in Each Performance Level for the Life Sciences Domain

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Below Standard** | **Near Standard** | **Above Standard** |
| Grade 5 | 150–199 | 200–233 | 234–250 |
| Grade 8 | 350–400 | 401–430 | 431–450 |
| High school | 550–598 | 599–629 | 630–650 |

Table 7.13 Scale Score Ranges in Each Performance Level for the Physical Sciences Domain

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Below Standard** | **Near Standard** | **Above Standard** |
| Grade 5 | 150–196 | 197–235 | 236–250 |
| Grade 8 | 350–400 | 401–430 | 431–450 |
| High school | 550–599 | 600–629 | 630–650 |

This science domain performance level estimation method for CAST is similar to, but slightly different from, the Smarter Balanced approach for a claim performance level. CAST uses the CSEM of the science domain at the level 3 theta threshold scores, while Smarter Balanced uses the CSEM of an individual student claim theta score.

Table 7.14 through table 7.16 show the percentages of science domain performance levels for Earth and Space Sciences, Life Sciences, and Physical Sciences, respectively. The adjusted domain scale cut scores now provide more information to classify students into level 1 through level 3, compared to the old domain cut scores. This is due to improved alignment with the blueprint and greater measurement precision of the items as item development, over time, has resulted in greater discrimination.

Table 7.14 Percentage of Students in Each Performance Level for the Earth and Space Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Number of Students Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| Grade 5 | 421,639 | 45.89 | 42.86 | 11.25 |
| Grade 8 | 426,195 | 39.85 | 48.36 | 11.79 |
| High school—Grade 10 | 24,835 | 54.42 | 35.37 | 10.22 |
| High school—Grade 11 | 331,021 | 47.88 | 37.38 | 14.75 |
| High school—Grade 12 | 99,215 | 54.23 | 32.77 | 13.00 |
| High school—All grades | 455,071 | 49.62 | 36.27 | 14.12 |

Table 7.15 Percentage of Students in Each Performance Level for the Life Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Number of Students Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| Grade 5 | 421,655 | 49.63 | 38.04 | 12.33 |
| Grade 8 | 426,134 | 51.74 | 36.20 | 12.05 |
| High school—Grade 10 | 24,809 | 50.18 | 38.61 | 11.22 |
| High school—Grade 11 | 330,785 | 44.59 | 39.62 | 15.79 |
| High school—Grade 12 | 99,118 | 51.29 | 35.48 | 13.24 |
| High school—All grades | 454,712 | 46.35 | 38.66 | 14.99 |

Table 7.16 Percentage of Students in Each Performance Level for the Physical Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Number of Students Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| Grade 5 | 421,625 | 47.86 | 42.40 | 9.74 |
| Grade 8 | 425,511 | 46.93 | 38.98 | 14.08 |
| High school—Grade 10 | 24,731 | 53.12 | 37.13 | 9.75 |
| High school—Grade 11 | 330,229 | 47.04 | 38.62 | 14.34 |
| High school—Grade 12 | 98,844 | 54.06 | 34.15 | 11.79 |
| High school—All grades | 453,804 | 48.90 | 37.57 | 13.54 |

Across all grade levels, the percentages of students in the Above Standard performance level ranged from 10.22 to 14.75 for Earth and Space Sciences, from 11.22 to 15.79 for Life Sciences, and from 9.74 to 14.34 for Physical Sciences. For all grade levels across science domains, 45.59 to 60.15 percent of students achieved either Near Standard or Above Standard levels.

The percentages of students identified as Near Standard and Above Standard were higher than the percentages of Below Standard for grades five and eight and the high school grade band; and science domains with the exception of the Life Sciences domain for grade eight. Note that only overall high school is discussed here, because the sample sizes for grades ten and twelve are relatively small.

Student group demographic summaries of science domain performance levels are presented in [appendix 7.E](#_Appendix_7.E:_Demographic) for grades five, eight, ten, eleven, and twelve, and for high school. Table 7.E.1 through table 7.E.6 report Earth and Space Sciences, table 7.E.7 through table 7.E.12 report Life Sciences, and table 7.E.13 through table 7.E.18 report Physical Sciences. The description of the demographic student groups included in these tables is shown in table 7.17.

#### Demographic Student Group Summaries

Demographic student groups included in the summaries in this technical report are shown in table 7.17. The number and the percentage of students for these demographic student groups are provided in [appendix 7.C](#_Appendix_7.C:_Demographic), starting in table 7.C.1 through table 7.C.5 for grades five, eight, ten, eleven, and twelve and in table 7.C.6 for high school. Additional information on specific demographic student groups can be found on the Test Results for California’s Assessments website.

Table 7.17 Demographic Student Groups to Be Reported

|  |  |
| --- | --- |
| **Category** | **Student Groups** |
| **Disability Status** | * Disability * No disability |
| **Economic Status** | * Economically disadvantaged * Not economically disadvantaged |
| **English Language Fluency** | * English learner (EL) * English only * Reclassified fluent English proficient (RFEP) * Initial fluent English proficient (IFEP) * Adult English learner (ADEL) * To be determined * English proficiency unknown |
| **Ethnicity** | * American Indian or Alaska Native * Asian * Native Hawaiian or Other Pacific Islander * Filipino * Hispanic or Latino * Black or African American * White * Two or more races |
| **Foster Youth Status** | * Foster youth * Not foster youth |
| **Gender** | * Male * Female * Nonbinary |
| **Homeless Status** | * Homeless * Not homeless |
| **Migrant Status** | * Eligible for the Title I Part C Migrant Program (Migrant education) * Not eligible for the Title I Part C Migrant Program (Not migrant education) |
| **Military Status** | * Armed forces family member * Not armed forces family member |

Demographic summaries, including the mean and SD of the overall scale score and the percentage of students in each of the achievement levels for total scores, are presented in table 7.D.1 through table 7.D.6 in [appendix 7.D](#_Appendix_7.D:_Demographic) for grades five, eight, ten, eleven, and twelve, as well as for the high school grade band. Refer to subsection [*7.3.1 Total-Test Achievement Levels*](#_Total-Test_Achievement_Levels) for a description of achievement levels.

The percentages of students who met or exceeded standards were similar between male and female students. Note that interpretation of the high percentages of the nonbinary groups is approached with caution because the sample sizes were small, ranging from 64 to 673 across all grade levels.

In terms of English proficiency, the IFEP group was among the highest performing, while the EL group was among the lowest performing. It is important to note that the sample sizes for the ADEL, to be determined, and unknown groups were small across all grade levels. For ethnicity, Asian, Filipino, White and student groups who identified as two or more races were among the highest performing.

The percentages of students who met or exceeded standards were similar between the two military status groups (armed forces family member and not armed forces family member). More students met or exceeded standards in the not economically disadvantaged, no disability, not migrant education, not homeless, and not foster youth student groups, compared to those in the economically disadvantaged, disability, migrant education, homeless, and foster youth student groups, respectively.

### Reports Produced and Scores for Each Report

The assessments that make up the CAASPP computer-based assessments provide results or score summaries that are reported for different purposes. The four major purposes are to

1. help facilitate conversations between parents/guardians and teachers about student performance,
2. serve as a tool to help parents/guardians and teachers work together to improve student learning,
3. help schools and LEAs identify strengths and areas that need improvement in their educational programs, and
4. provide the public and policymakers with information about student achievement.

This section provides detailed descriptions of the uses and applications of CAASPP reporting for students.

#### Online Reporting

TOMS is a secure website hosted by ETS that permits LEA users to manage the CAASPP computer-based assessments and to inform the TDS. This system uses a role-specific design to restrict access to certain tools and applications based on the user’s designated role. Specific functions of TOMS include the following:

* Manage user access privileges
* Manage test administration calendars and testing windows
* Manage student test assignments
* Manage and confirm the accuracy of students’ test settings (that is, designated supports and accommodations) prior to testing
* Generate and download various reports

In addition to TOMS, another California online reporting system was used during the 2023–‍24 administration: the California Educator Reporting System (CERS).

TOMS communicated with CERS, which provided authorized users with interactive and cumulative online reports for science domains at the student, school, group, and LEA levels. CERS provided educators and administrators access to student results for each administered assessment available in the reporting system.

Based on the CAASPP reporting requirements, CERS provided the summative reports containing information outlining student knowledge and skills as they became available. CERS also permitted access to individual score reports, which provided score data for each administered assessment available in the reporting system. The online aggregate reports were available to be downloaded in PDF, Excel, and comma-separated value formats.

CERS was the primary source for LEA staff to analyze CAASPP results at the LEA, school, grade, classroom, or customized group level. CERS provided these reports, which can be downloaded and used to inform instruction. LEA staff with TOMS logon credentials could enter CERS through the CAASPP & English Language Proficiency Assessments for California (ELPAC) Website to access student assessment results.

#### Special Cases

Student scores were not reported for the following cases:

* The student had a medical emergency during testing.
* The student’s parent/guardian requested exemption from testing.
* The student did not log on to test systems.
* The student score was invalidated in the system (not reported in aggregated reporting).

#### Types of Score Reports

There are four categories of CAASPP reports. Reports within each category are presented in this subsection.

1. **Student Score Report (SSR)—**The SSR was the official score report for parents/‌guardians. An SSR described the student’s results.
2. **LEA student data files—**LEA student data files were available for download on demand by the LEA in TOMS to coincide with availability of the SSRs.
3. **LEA aggregations—**Aggregate data was available to view in CERS and the Test Results for California’s Assessments website.
4. **CERS reports—**CERS allows users to produce customizable reports for students individually and on the aggregate. Individual reports are based on selectable criteria that includes the school year, content area, and assessment type; these reports include the student’s scale score and error band information, reporting category, and date of the assessment. CERS aggregate reports allow customization based on such criteria as student groups, content areas, and year tested.

##### Student Score Report

The CAST SSR is the official score report for parents/guardians and includes the following metrics:

* Reported scale score (The ranges of scale scores are provided in table 7.8.)
* Reported achievement levels (CAST achievement levels are “Standard Exceeded,” “Standard Met,” “Standard Nearly Met,” and “Standard Not Met.”)
* Reported science-domain-specific performance levels (The science-domain-specific performance levels are “Above Standard,” “Near Standard,” and “Below Standard.”)

LEAs had three options for accessing and distributing SSRs to parents/guardians:

1. Accessing electronic SSRs in PDF or HTML format using a locally provided parent/guardian or student portal
2. Downloading SSRs in PDF or HTML format from TOMS and making them available electronically using a secure local method
3. Downloading SSR PDFs from TOMS, printing them, and making them available locally

The LEA CAASPP coordinator could securely deliver the appropriate reports to test sites. In the case of a locally printed CAST SSR, the LEA sent the printed report(s) to the child’s parent/guardian. CAST SSRs that included individual student results were not distributed beyond the student’s school.

Scores for students who were assigned accommodations or designated supports are reported in the same way as for students who were not assigned accommodations or designated supports. Detailed information about accessibility resources is described in subsection [*5.4.1 Accessibility Resource Categories*](#_Accessibility_Resource_Categories_1).

For the 2023–24 test administration, SSRs were made available to the LEAs in multiple supported languages, including English, Spanish, Arabic, Chinese (Traditional), Filipino, Korean, and Vietnamese. An SSR in a supported language was created if the student’s primary language as reported in the California Longitudinal Pupil Achievement Data System was one of these supported languages. The LEAs that received SSRs in supported languages received one SSR in English and another in the supported language. These reports were available as PDFs for the LEA to download from TOMS.

Further information about the SSR and its interpretation is provided on the CAASPP Starting Smarter website for California assessments.

###### Access via Student or Parent Portal

LEAs had the option to provide SSRs electronically using a locally provided parent or student portal.

Amazon Web Services—with the Amazon Simple Storage Service and the Amazon Key Management Service—ensured encrypted access for parents/guardians to view a child’s electronic SSR, which was available as a PDF.

###### Access via the Test Operations Management System

The LEA CAASPP coordinator downloaded the electronic PDFs directly from TOMS and could securely deliver the appropriate reports to test sites. Optionally, the LEA could download the SSR PDF, print the SSR PDF, and then provide the printed report(s) to the child’s parent/‌guardian.

##### Local Educational Agency Student Data Files and Aggregations

The CAASPP student data files for the LEA were available for the LEA CAASPP coordinator and site CAASPP coordinator to download from TOMS.

Student scores and aggregations were also available to LEAs prior to the release of final reports via electronic reporting, using CERS. This website permitted LEAs to view preliminary results data for all assessments taken as it became available.

Current and historical aggregate results are accessible to the public on the CDE Test Results for California’s Assessments website.

##### California Educator Reporting System Reports

CERS was the primary source for LEA staff to analyze CAASPP results at the LEA, school, grade, classroom, or customized group level. CERS provided these reports, which can be downloaded and used to inform instruction. LEA staff with TOMS logon credentials could enter CERS through the CAASPP & ELPAC Website to access student assessment results.

There are four additional report types in addition to the Individual Student Reports:

1. **Yearly Reports** are the basic performance report. These reports summarize summative assessment performance for student populations from one or more grade levels for one or more years of available data.
2. **Longitudinal Reports** track summative assessment performance for a single student population as that population progresses through different grades. In addition to presenting tabular data, these reports include a line graph showing how the performance of the population changed from grade to grade. These reports are only available if the reporting system includes summative assessment results.
3. **Claim and Domain Reports** are yearly reports that break down performance data by claim or domain for summative assessments.
4. **Target Reports** are yearly reports that break down performance by claim and target for the Smarter Balanced summative assessments.

#### Score Report Applications

CAST results, presented in SSRs, provided parents/guardians with information about their child’s progress. The results were one tool for increasing communication and collaboration between parents/guardians and teachers about ways to improve their child’s achievement of the CA NGSS. They provided limited information about one measure of a student’s academic performance. Like any important measure of student performance, the test results should be viewed with other available information such as progress on individualized education program goals, assignments, and teacher conferences.

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

CAST results helped schools and LEAs identify strengths and areas of growth in their instructional programs. Each year, staff from schools and LEAs examine CAST results at each grade level 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 the needs of students, and
* decisions about how to use funds to ensure that students achieve the standards.

#### Criteria for Interpreting Test Scores

An LEA may use CAASPP computer-based summative assessment 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 assessment can provide only limited information. Other relevant information should be considered as well. It is advisable for parents/guardians 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 CAASPP computer-based summative assessment results. It is also important to note that a student’s score in a content area 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 achievement-level results, the user is limited to comparisons within a grade level or grade band. The user may compare scale scores for the same grade level, within a school, between schools, or between a school and its LEA, its county, or the state. The CAASPP user can also make comparisons within the same grade level or grade band across years.

However, comparing scale scores from different grade levels for the CAASPP is not appropriate, because the curricula are different across grade levels and the scale scores are not vertically linked between grade levels.

For more details on the criteria for interpreting information provided on the score reports, refer to the CAASPP Starting Smarter website for California assessmentsor the *2023–24 CAASPP and ELPAC Scoring and Reporting Guide* (CDE, 2024), which was applicable for the 2023–24 CAASPP administration.

### References

California Department of Education. (2024). *CAASPP and ELPAC scoring and reporting guide*. Sacramento, CA: California Department of Education.

Drucker, Harris, Burges, Christopher J. C., Kaufman, Linda, Smola, Alexander J., & Vapnik, Vladimir N. (1996). Support vector regression machines. In *Advances in neural information processing systems 9*, NIPS 1996 (pp. 155–61). Cambridge, MA: MIT Press.

Stocking, M. L. (1996). *An alternative method for scoring adaptive tests*. Journal of Educational and Behavioral Statistics, 21, 365–89.

Vapnik, Vladimir N. (1995). *The nature of statistical learning theory.* New York, NY: Springer-Verlag.

### Accessibility Information

#### Alternative Text for Equation 7.1

Number-right sub j is equal to n times the fraction with the numerator exponent based theta-hat sub j and the denominator one plus exponent theta-hat sub j. *(Return to* [*equation 7.1*](#EQ7_1)*.)*

#### Alternative Text for Equation 7.2

Below Standard level is equal to a science domain theta score that is less than the level 3 theta threshold score of the total assessment minus open parentheses one decimal point five multiplied by the conditional standard error of the science domain theta scores close parentheses. *(Return to* [*equation 7.2*](#EQ7_2)*.)*

#### Alternative Text for Equation 7.3

Near Standard level is equal to the level 3 theta threshold score of the total assessment minus open parentheses one decimal point five multiplied by the conditional standard error of the science domain theta scores close parentheses that is less than or equal to a science domain theta score that is less than the level 3 theta threshold score of the total assessment plus open parentheses one decimal point five multiplied by the conditional standard error of the science domain theta scores close parentheses. *(Return to* [*equation 7.3*](#EQ7_3)*.)*

#### Alternative Text for Equation 7.4

Above Standard level is equal to a science domain theta score that is greater than or equal to the level 3 theta threshold score of the total assessment plus open parentheses one decimal point five multiplied by the conditional standard error of the science domain theta scores close parentheses. *(Return to* [*equation 7.4*](#EQ7_4)*.)*

### Appendix 7.A: Overall Theta Score Distribution

Table 7.A.1 Overall Theta Score Distribution for Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Theta Score** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| θ = −4.0 | 3 | <0.1 | 3 | <0.1 |
| −4.0 < θ ≤ −3.5 | 36 | <0.1 | 39 | <0.1 |
| −3.5 < θ ≤ −3.0 | 140 | <0.1 | 179 | <0.1 |
| −3.0 < θ ≤ −2.5 | 631 | 0.1 | 810 | 0.2 |
| −2.5 < θ ≤ −2.0 | 5,376 | 1.3 | 6,186 | 1.5 |
| −2.0 < θ ≤ −1.5 | 19,759 | 4.7 | 25,945 | 6.1 |
| −1.5 < θ ≤ −1.0 | 48,614 | 11.5 | 74,559 | 17.7 |
| −1.0 < θ ≤ −0.5 | 69,092 | 16.4 | 143,651 | 34.0 |
| −0.5 < θ ≤ 0 | 74,987 | 17.8 | 218,638 | 51.8 |
| 0 < θ ≤ 0.5 | 68,439 | 16.2 | 287,077 | 68.0 |
| 0.5 < θ ≤ 1.0 | 54,522 | 12.9 | 341,599 | 80.9 |
| 1.0 < θ ≤ 1.5 | 40,695 | 9.6 | 382,294 | 90.5 |
| 1.5 < θ ≤ 2.0 | 21,663 | 5.1 | 403,957 | 95.6 |
| 2.0 < θ ≤ 2.5 | 10,369 | 2.5 | 414,326 | 98.1 |
| 2.5 < θ ≤ 3.0 | 5,316 | 1.3 | 419,642 | 99.3 |
| 3.0 < θ ≤ 3.5 | 1,142 | 0.3 | 420,784 | 99.6 |
| 3.5 < θ < 4.0 | 1,066 | 0.3 | 421,850 | 99.9 |
| θ = 4.0 | 548 | 0.1 | 422,398 | 100.0 |

Table 7.A.2 Overall Theta Score Distribution for Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Theta Score** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| θ = −4.0 | 99 | <0.1 | 99 | <0.1 |
| −4.0 < θ ≤ −3.5 | 0 | 0.0 | 99 | <0.1 |
| −3.5 < θ ≤ −3.0 | 322 | <0.1 | 421 | <0.1 |
| −3.0 < θ ≤ −2.5 | 1,122 | 0.3 | 1,543 | 0.4 |
| −2.5 < θ ≤ −2.0 | 2,745 | 0.6 | 4,288 | 1.0 |
| −2.0 < θ ≤ −1.5 | 19,379 | 4.5 | 23,667 | 5.5 |
| −1.5 < θ ≤ −1.0 | 53,901 | 12.6 | 77,568 | 18.1 |
| −1.0 < θ ≤ −0.5 | 73,952 | 17.3 | 151,520 | 35.4 |
| −0.5 < θ ≤ 0 | 76,972 | 18.0 | 228,492 | 53.5 |
| 0 < θ ≤ 0.5 | 69,944 | 16.4 | 298,436 | 69.8 |
| 0.5 < θ ≤ 1.0 | 53,624 | 12.5 | 352,060 | 82.4 |
| 1.0 < θ ≤ 1.5 | 35,068 | 8.2 | 387,128 | 90.6 |
| 1.5 < θ ≤ 2.0 | 20,151 | 4.7 | 407,279 | 95.3 |
| 2.0 < θ ≤ 2.5 | 10,447 | 2.4 | 417,726 | 97.7 |
| 2.5 < θ ≤ 3.0 | 4,986 | 1.2 | 422,712 | 98.9 |
| 3.0 < θ ≤ 3.5 | 2,317 | 0.5 | 425,029 | 99.4 |
| 3.5 < θ < 4.0 | 1,269 | 0.3 | 426,298 | 99.7 |
| θ = 4.0 | 1,150 | 0.3 | 427,448 | 100.0 |

Table 7.A.3 Overall Theta Score Distribution for Grade Ten

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Theta Score** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| θ = −4.0 | 1 | <0.1 | 1 | <0.1 |
| −4.0 < θ ≤ −3.5 | 1 | <0.1 | 2 | <0.1 |
| −3.5 < θ ≤ −3.0 | 1 | <0.1 | 3 | <0.1 |
| −3.0 < θ ≤ −2.5 | 16 | <0.1 | 19 | <0.1 |
| −2.5 < θ ≤ −2.0 | 76 | 0.3 | 95 | 0.4 |
| −2.0 < θ ≤ −1.5 | 758 | 3.0 | 853 | 3.4 |
| −1.5 < θ ≤ −1.0 | 3,466 | 13.9 | 4,319 | 17.3 |
| −1.0 < θ ≤ −0.5 | 5,317 | 21.3 | 9,636 | 38.6 |
| −0.5 < θ ≤ 0 | 4,908 | 19.7 | 14,544 | 58.3 |
| 0 < θ ≤ 0.5 | 4,193 | 16.8 | 18,737 | 75.1 |
| 0.5 < θ ≤ 1.0 | 2,866 | 11.5 | 21,603 | 86.6 |
| 1.0 < θ ≤ 1.5 | 1,741 | 7.0 | 23,344 | 93.6 |
| 1.5 < θ ≤ 2.0 | 938 | 3.8 | 24,282 | 97.4 |
| 2.0 < θ ≤ 2.5 | 402 | 1.6 | 24,684 | 99.0 |
| 2.5 < θ ≤ 3.0 | 150 | 0.6 | 24,834 | 99.6 |
| 3.0 < θ ≤ 3.5 | 58 | 0.2 | 24,892 | 99.8 |
| 3.5 < θ < 4.0 | 22 | <0.1 | 24,914 | 99.9 |
| θ = 4.0 | 21 | <0.1 | 24,935 | 100.0 |

Table 7.A.4 Overall Theta Score Distribution for Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Theta Score** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| θ = −4.0 | 7 | <0.1 | 7 | <0.1 |
| −4.0 < θ ≤ −3.5 | 7 | <0.1 | 14 | <0.1 |
| −3.5 < θ ≤ −3.0 | 17 | <0.1 | 31 | <0.1 |
| −3.0 < θ ≤ −2.5 | 185 | <0.1 | 216 | <0.1 |
| −2.5 < θ ≤ −2.0 | 1,101 | 0.3 | 1,317 | 0.4 |
| −2.0 < θ ≤ −1.5 | 10,226 | 3.1 | 11,543 | 3.5 |
| −1.5 < θ ≤ −1.0 | 40,899 | 12.3 | 52,442 | 15.8 |
| −1.0 < θ ≤ −0.5 | 59,943 | 18.1 | 112,385 | 33.9 |
| −0.5 < θ ≤ 0 | 58,063 | 17.5 | 170,448 | 51.3 |
| 0 < θ ≤ 0.5 | 54,567 | 16.4 | 225,015 | 67.8 |
| 0.5 < θ ≤ 1.0 | 43,427 | 13.1 | 268,442 | 80.9 |
| 1.0 < θ ≤ 1.5 | 30,030 | 9.0 | 298,472 | 89.9 |
| 1.5 < θ ≤ 2.0 | 18,609 | 5.6 | 317,081 | 95.5 |
| 2.0 < θ ≤ 2.5 | 8,913 | 2.7 | 325,994 | 98.2 |
| 2.5 < θ ≤ 3.0 | 3,598 | 1.1 | 329,592 | 99.3 |
| 3.0 < θ ≤ 3.5 | 1,366 | 0.4 | 330,958 | 99.7 |
| 3.5 < θ < 4.0 | 524 | 0.2 | 331,482 | 99.9 |
| θ = 4.0 | 467 | 0.1 | 331,949 | 100.0 |

Table 7.A.5 Overall Theta Score Distribution for Grade Twelve

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Theta Score** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| θ = −4.0 | 6 | <0.1 | 6 | <0.1 |
| −4.0 < θ ≤ −3.5 | 3 | <0.1 | 9 | <0.1 |
| −3.5 < θ ≤ −3.0 | 8 | <0.1 | 17 | <0.1 |
| −3.0 < θ ≤ −2.5 | 72 | <0.1 | 89 | <0.1 |
| −2.5 < θ ≤ −2.0 | 501 | 0.5 | 590 | 0.6 |
| −2.0 < θ ≤ −1.5 | 4,320 | 4.3 | 4,910 | 4.9 |
| −1.5 < θ ≤ −1.0 | 15,327 | 15.4 | 20,237 | 20.3 |
| −1.0 < θ ≤ −0.5 | 20,293 | 20.4 | 40,530 | 40.7 |
| −0.5 < θ ≤ 0 | 17,586 | 17.7 | 58,116 | 58.3 |
| 0 < θ ≤ 0.5 | 14,644 | 14.7 | 72,760 | 73.1 |
| 0.5 < θ ≤ 1.0 | 10,802 | 10.8 | 83,562 | 83.9 |
| 1.0 < θ ≤ 1.5 | 7,394 | 7.4 | 90,956 | 91.3 |
| 1.5 < θ ≤ 2.0 | 4,539 | 4.6 | 95,495 | 95.9 |
| 2.0 < θ ≤ 2.5 | 2,393 | 2.4 | 97,888 | 98.3 |
| 2.5 < θ ≤ 3.0 | 1,021 | 1.0 | 98,909 | 99.3 |
| 3.0 < θ ≤ 3.5 | 395 | 0.4 | 99,304 | 99.7 |
| 3.5 < θ < 4.0 | 160 | 0.2 | 99,464 | 99.9 |
| θ = 4.0 | 138 | 0.1 | 99,602 | 100.0 |

Table 7.A.6 Overall Theta Score Distribution for High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Theta Score** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| θ = −4.0 | 14 | <0.1 | 14 | <0.1 |
| −4.0 < θ ≤ −3.5 | 11 | <0.1 | 25 | <0.1 |
| −3.5 < θ ≤ −3.0 | 26 | <0.1 | 51 | <0.1 |
| −3.0 < θ ≤ −2.5 | 273 | <0.1 | 324 | <0.1 |
| −2.5 < θ ≤ −2.0 | 1,678 | 0.4 | 2,002 | 0.4 |
| −2.0 < θ ≤ −1.5 | 15,304 | 3.4 | 17,306 | 3.8 |
| −1.5 < θ ≤ −1.0 | 59,692 | 13.1 | 76,998 | 16.9 |
| −1.0 < θ ≤ −0.5 | 85,553 | 18.7 | 162,551 | 35.6 |
| −0.5 < θ ≤ 0 | 80,557 | 17.6 | 243,108 | 53.3 |
| 0 < θ ≤ 0.5 | 73,404 | 16.1 | 316,512 | 69.3 |
| 0.5 < θ ≤ 1.0 | 57,095 | 12.5 | 373,607 | 81.8 |
| 1.0 < θ ≤ 1.5 | 39,165 | 8.6 | 412,772 | 90.4 |
| 1.5 < θ ≤ 2.0 | 24,086 | 5.3 | 436,858 | 95.7 |
| 2.0 < θ ≤ 2.5 | 11,708 | 2.6 | 448,566 | 98.3 |
| 2.5 < θ ≤ 3.0 | 4,769 | 1.0 | 453,335 | 99.3 |
| 3.0 < θ ≤ 3.5 | 1,819 | 0.4 | 455,154 | 99.7 |
| 3.5 < θ < 4.0 | 706 | 0.2 | 455,860 | 99.9 |
| θ = 4.0 | 626 | 0.1 | 456,486 | 100.0 |

### Appendix 7.B: Scale Scores of Assessments

Table 7.B.1 CAST Grade Five Scale Scores

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scale Score** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 150 | 122 | <0.1 | 122 | <0.1 |
| 151 | 1 | <0.1 | 123 | <0.1 |
| 152 | 9 | <0.1 | 132 | <0.1 |
| 153 | 22 | <0.1 | 154 | <0.1 |
| 154 | 31 | <0.1 | 185 | <0.1 |
| 155 | 149 | <0.1 | 334 | <0.1 |
| 156 | 17 | <0.1 | 351 | <0.1 |
| 157 | 461 | 0.1 | 812 | 0.2 |
| 158 | 18 | <0.1 | 830 | 0.2 |
| 159 | 970 | 0.2 | 1,800 | 0.4 |
| 160 | 458 | 0.1 | 2,258 | 0.5 |
| 161 | 1,411 | 0.3 | 3,669 | 0.9 |
| 162 | 1,611 | 0.4 | 5,280 | 1.2 |
| 163 | 1,618 | 0.4 | 6,898 | 1.6 |
| 164 | 2,327 | 0.6 | 9,225 | 2.2 |
| 165 | 2,456 | 0.6 | 11,681 | 2.8 |
| 166 | 3,059 | 0.7 | 14,740 | 3.5 |
| 167 | 3,279 | 0.8 | 18,019 | 4.3 |
| 168 | 4,111 | 1.0 | 22,130 | 5.2 |
| 169 | 4,227 | 1.0 | 26,357 | 6.2 |
| 170 | 4,770 | 1.1 | 31,127 | 7.4 |
| 171 | 4,921 | 1.2 | 36,048 | 8.5 |
| 172 | 5,307 | 1.3 | 41,355 | 9.8 |
| 173 | 5,610 | 1.3 | 46,965 | 11.1 |
| 174 | 5,713 | 1.4 | 52,678 | 12.5 |
| 175 | 5,914 | 1.4 | 58,592 | 13.9 |
| 176 | 7,389 | 1.7 | 65,981 | 15.6 |
| 177 | 4,596 | 1.1 | 70,577 | 16.7 |
| 178 | 7,595 | 1.8 | 78,172 | 18.5 |
| 179 | 4,822 | 1.1 | 82,994 | 19.6 |
| 180 | 9,233 | 2.2 | 92,227 | 21.8 |
| 181 | 3,138 | 0.7 | 95,365 | 22.6 |
| 182 | 9,137 | 2.2 | 104,502 | 24.7 |
| 183 | 3,123 | 0.7 | 107,625 | 25.5 |
| 184 | 9,083 | 2.1 | 116,708 | 27.6 |
| 185 | 3,180 | 0.8 | 119,888 | 28.4 |
| 186 | 10,623 | 2.5 | 130,511 | 30.9 |
| 187 | 3,045 | 0.7 | 133,556 | 31.6 |
| 188 | 8,965 | 2.1 | 142,521 | 33.7 |
| 189 | 2,986 | 0.7 | 145,507 | 34.4 |
| 190 | 9,011 | 2.1 | 154,518 | 36.6 |
| 191 | 4,393 | 1.0 | 158,911 | 37.6 |
| 192 | 8,931 | 2.1 | 167,842 | 39.7 |
| 193 | 2,858 | 0.7 | 170,700 | 40.4 |
| 194 | 8,893 | 2.1 | 179,593 | 42.5 |
| 195 | 4,399 | 1.0 | 183,992 | 43.5 |
| 196 | 5,929 | 1.4 | 189,921 | 44.9 |
| 197 | 5,893 | 1.4 | 195,814 | 46.3 |
| 198 | 5,853 | 1.4 | 201,667 | 47.7 |
| 199 | 5,808 | 1.4 | 207,475 | 49.1 |
| 200 | 4,324 | 1.0 | 211,799 | 50.1 |
| 201 | 7,176 | 1.7 | 218,975 | 51.8 |
| 202 | 4,365 | 1.0 | 223,340 | 52.9 |
| 203 | 7,420 | 1.8 | 230,760 | 54.6 |
| 204 | 4,383 | 1.0 | 235,143 | 55.7 |
| 205 | 7,367 | 1.7 | 242,510 | 57.4 |
| 206 | 4,351 | 1.0 | 246,861 | 58.4 |
| 207 | 5,758 | 1.4 | 252,619 | 59.8 |
| 208 | 5,882 | 1.4 | 258,501 | 61.2 |
| 209 | 5,750 | 1.4 | 264,251 | 62.5 |
| 210 | 5,774 | 1.4 | 270,025 | 63.9 |
| 211 | 5,842 | 1.4 | 275,867 | 65.3 |
| 212 | 5,843 | 1.4 | 281,710 | 66.7 |
| 213 | 2,762 | 0.7 | 284,472 | 67.3 |
| 214 | 8,456 | 2.0 | 292,928 | 69.3 |
| 215 | 2,782 | 0.7 | 295,710 | 70.0 |
| 216 | 7,014 | 1.7 | 302,724 | 71.6 |
| 217 | 4,358 | 1.0 | 307,082 | 72.7 |
| 218 | 5,653 | 1.3 | 312,735 | 74.0 |
| 219 | 5,615 | 1.3 | 318,350 | 75.3 |
| 220 | 4,160 | 1.0 | 322,510 | 76.3 |
| 221 | 7,063 | 1.7 | 329,573 | 78.0 |
| 222 | 1,283 | 0.3 | 330,856 | 78.3 |
| 223 | 9,627 | 2.3 | 340,483 | 80.6 |
| 224 | 1,293 | 0.3 | 341,776 | 80.9 |
| 225 | 8,032 | 1.9 | 349,808 | 82.8 |
| 226 | 1,342 | 0.3 | 351,150 | 83.1 |
| 227 | 7,525 | 1.8 | 358,675 | 84.9 |
| 228 | 2,676 | 0.6 | 361,351 | 85.5 |
| 229 | 5,920 | 1.4 | 367,271 | 86.9 |
| 230 | 3,795 | 0.9 | 371,066 | 87.8 |
| 231 | 3,394 | 0.8 | 374,460 | 88.6 |
| 232 | 4,653 | 1.1 | 379,113 | 89.7 |
| 233 | 3,328 | 0.8 | 382,441 | 90.5 |
| 234 | 4,205 | 1.0 | 386,646 | 91.5 |
| 235 | 4,220 | 1.0 | 390,866 | 92.5 |
| 236 | 3,819 | 0.9 | 394,685 | 93.4 |
| 237 | 3,702 | 0.9 | 398,387 | 94.3 |
| 238 | 2,590 | 0.6 | 400,977 | 94.9 |
| 239 | 3,109 | 0.7 | 404,086 | 95.6 |
| 240 | 3,013 | 0.7 | 407,099 | 96.4 |
| 241 | 2,029 | 0.5 | 409,128 | 96.8 |
| 242 | 2,447 | 0.6 | 411,575 | 97.4 |
| 243 | 2,370 | 0.6 | 413,945 | 98.0 |
| 244 | 1,432 | 0.3 | 415,377 | 98.3 |
| 245 | 2,210 | 0.5 | 417,587 | 98.8 |
| 246 | 1,366 | 0.3 | 418,953 | 99.2 |
| 247 | 1,103 | 0.3 | 420,056 | 99.4 |
| 248 | 850 | 0.2 | 420,906 | 99.6 |
| 249 | 642 | 0.2 | 421,548 | 99.8 |
| 250 | 972 | 0.2 | 422,520 | 100.0 |

Table 7.B.2 CAST Grade Eight Scale Scores

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scale Score** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 350 | 504 | 0.1 | 504 | 0.1 |
| 351 | 6 | <0.1 | 510 | 0.1 |
| 352 | 11 | <0.1 | 521 | 0.1 |
| 353 | 118 | <0.1 | 639 | 0.1 |
| 354 | 9 | <0.1 | 648 | 0.2 |
| 355 | 327 | <0.1 | 975 | 0.2 |
| 356 | 93 | <0.1 | 1,068 | 0.2 |
| 357 | 186 | <0.1 | 1,254 | 0.3 |
| 358 | 828 | 0.2 | 2,082 | 0.5 |
| 359 | 270 | <0.1 | 2,352 | 0.5 |
| 360 | 1,196 | 0.3 | 3,548 | 0.8 |
| 361 | 368 | <0.1 | 3,916 | 0.9 |
| 362 | 1,340 | 0.3 | 5,256 | 1.2 |
| 363 | 2,338 | 0.5 | 7,594 | 1.8 |
| 364 | 698 | 0.2 | 8,292 | 1.9 |
| 365 | 2,469 | 0.6 | 10,761 | 2.5 |
| 366 | 3,838 | 0.9 | 14,599 | 3.4 |
| 367 | 1,067 | 0.2 | 15,666 | 3.7 |
| 368 | 5,103 | 1.2 | 20,769 | 4.9 |
| 369 | 4,233 | 1.0 | 25,002 | 5.8 |
| 370 | 3,074 | 0.7 | 28,076 | 6.6 |
| 371 | 10,239 | 2.4 | 38,315 | 9.0 |
| 372 | 42 | <0.1 | 38,357 | 9.0 |
| 373 | 11,683 | 2.7 | 50,040 | 11.7 |
| 374 | 3,982 | 0.9 | 54,022 | 12.6 |
| 375 | 4,351 | 1.0 | 58,373 | 13.6 |
| 376 | 12,437 | 2.9 | 70,810 | 16.5 |
| 377 | 57 | <0.1 | 70,867 | 16.6 |
| 378 | 10,150 | 2.4 | 81,017 | 18.9 |
| 379 | 6,382 | 1.5 | 87,399 | 20.4 |
| 380 | 7,983 | 1.9 | 95,382 | 22.3 |
| 381 | 8,012 | 1.9 | 103,394 | 24.2 |
| 382 | 5,272 | 1.2 | 108,666 | 25.4 |
| 383 | 7,739 | 1.8 | 116,405 | 27.2 |
| 384 | 5,524 | 1.3 | 121,929 | 28.5 |
| 385 | 7,013 | 1.6 | 128,942 | 30.1 |
| 386 | 7,107 | 1.7 | 136,049 | 31.8 |
| 387 | 6,988 | 1.6 | 143,037 | 33.4 |
| 388 | 6,705 | 1.6 | 149,742 | 35.0 |
| 389 | 6,575 | 1.5 | 156,317 | 36.5 |
| 390 | 4,831 | 1.1 | 161,148 | 37.7 |
| 391 | 6,402 | 1.5 | 167,550 | 39.2 |
| 392 | 7,750 | 1.8 | 175,300 | 41.0 |
| 393 | 6,090 | 1.4 | 181,390 | 42.4 |
| 394 | 6,065 | 1.4 | 187,455 | 43.8 |
| 395 | 6,113 | 1.4 | 193,568 | 45.2 |
| 396 | 5,992 | 1.4 | 199,560 | 46.6 |
| 397 | 5,864 | 1.4 | 205,424 | 48.0 |
| 398 | 5,771 | 1.3 | 211,195 | 49.4 |
| 399 | 5,660 | 1.3 | 216,855 | 50.7 |
| 400 | 5,722 | 1.3 | 222,577 | 52.0 |
| 401 | 8,335 | 1.9 | 230,912 | 54.0 |
| 402 | 2,840 | 0.7 | 233,752 | 54.6 |
| 403 | 8,119 | 1.9 | 241,871 | 56.5 |
| 404 | 5,497 | 1.3 | 247,368 | 57.8 |
| 405 | 5,492 | 1.3 | 252,860 | 59.1 |
| 406 | 6,799 | 1.6 | 259,659 | 60.7 |
| 407 | 3,828 | 0.9 | 263,487 | 61.6 |
| 408 | 9,052 | 2.1 | 272,539 | 63.7 |
| 409 | 1,296 | 0.3 | 273,835 | 64.0 |
| 410 | 10,409 | 2.4 | 284,244 | 66.4 |
| 411 | 1,238 | 0.3 | 285,482 | 66.7 |
| 412 | 8,874 | 2.1 | 294,356 | 68.8 |
| 413 | 2,504 | 0.6 | 296,860 | 69.4 |
| 414 | 7,363 | 1.7 | 304,223 | 71.1 |
| 415 | 4,739 | 1.1 | 308,962 | 72.2 |
| 416 | 4,805 | 1.1 | 313,767 | 73.3 |
| 417 | 4,655 | 1.1 | 318,422 | 74.4 |
| 418 | 4,738 | 1.1 | 323,160 | 75.5 |
| 419 | 6,795 | 1.6 | 329,955 | 77.1 |
| 420 | 2,353 | 0.5 | 332,308 | 77.7 |
| 421 | 8,905 | 2.1 | 341,213 | 79.7 |
| 422 | 7 | <0.1 | 341,220 | 79.7 |
| 423 | 8,470 | 2.0 | 349,690 | 81.7 |
| 424 | 1,983 | 0.5 | 351,673 | 82.2 |
| 425 | 6,552 | 1.5 | 358,225 | 83.7 |
| 426 | 2,901 | 0.7 | 361,126 | 84.4 |
| 427 | 4,980 | 1.2 | 366,106 | 85.5 |
| 428 | 2,694 | 0.6 | 368,800 | 86.2 |
| 429 | 4,929 | 1.2 | 373,729 | 87.3 |
| 430 | 4,453 | 1.0 | 378,182 | 88.4 |
| 431 | 3,670 | 0.9 | 381,852 | 89.2 |
| 432 | 3,313 | 0.8 | 385,165 | 90.0 |
| 433 | 3,351 | 0.8 | 388,516 | 90.8 |
| 434 | 3,143 | 0.7 | 391,659 | 91.5 |
| 435 | 3,117 | 0.7 | 394,776 | 92.2 |
| 436 | 3,676 | 0.9 | 398,452 | 93.1 |
| 437 | 2,883 | 0.7 | 401,335 | 93.8 |
| 438 | 2,694 | 0.6 | 404,029 | 94.4 |
| 439 | 2,578 | 0.6 | 406,607 | 95.0 |
| 440 | 3,558 | 0.8 | 410,165 | 95.8 |
| 441 | 2,266 | 0.5 | 412,431 | 96.4 |
| 442 | 2,073 | 0.5 | 414,504 | 96.9 |
| 443 | 2,389 | 0.6 | 416,893 | 97.4 |
| 444 | 2,158 | 0.5 | 419,051 | 97.9 |
| 445 | 1,523 | 0.4 | 420,574 | 98.3 |
| 446 | 1,777 | 0.4 | 422,351 | 98.7 |
| 447 | 1,382 | 0.3 | 423,733 | 99.0 |
| 448 | 1,440 | 0.3 | 425,173 | 99.4 |
| 449 | 1,066 | 0.2 | 426,239 | 99.6 |
| 450 | 1,713 | 0.4 | 427,952 | 100.0 |

Table 7.B.3 CAST Grade Ten Scale Scores

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scale Score** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 550 | 37 | 0.1 | 37 | 0.1 |
| 551 | 0 | 0.0 | 37 | 0.1 |
| 552 | 0 | 0.0 | 37 | 0.1 |
| 553 | 1 | <0.1 | 38 | 0.2 |
| 554 | 3 | <0.1 | 41 | 0.2 |
| 555 | 2 | <0.1 | 43 | 0.2 |
| 556 | 14 | <0.1 | 57 | 0.2 |
| 557 | 3 | <0.1 | 60 | 0.2 |
| 558 | 11 | <0.1 | 71 | 0.3 |
| 559 | 22 | <0.1 | 93 | 0.4 |
| 560 | 5 | <0.1 | 98 | 0.4 |
| 561 | 59 | 0.2 | 157 | 0.6 |
| 562 | 8 | <0.1 | 165 | 0.7 |
| 563 | 73 | 0.3 | 238 | 1.0 |
| 564 | 99 | 0.4 | 337 | 1.3 |
| 565 | 28 | 0.1 | 365 | 1.5 |
| 566 | 227 | 0.9 | 592 | 2.4 |
| 567 | 2 | <0.1 | 594 | 2.4 |
| 568 | 212 | 0.8 | 806 | 3.2 |
| 569 | 155 | 0.6 | 961 | 3.8 |
| 570 | 285 | 1.1 | 1,246 | 5.0 |
| 571 | 321 | 1.3 | 1,567 | 6.3 |
| 572 | 302 | 1.2 | 1,869 | 7.5 |
| 573 | 449 | 1.8 | 2,318 | 9.3 |
| 574 | 330 | 1.3 | 2,648 | 10.6 |
| 575 | 313 | 1.3 | 2,961 | 11.9 |
| 576 | 482 | 1.9 | 3,443 | 13.8 |
| 577 | 473 | 1.9 | 3,916 | 15.7 |
| 578 | 528 | 2.1 | 4,444 | 17.8 |
| 579 | 512 | 2.1 | 4,956 | 19.8 |
| 580 | 520 | 2.1 | 5,476 | 21.9 |
| 581 | 521 | 2.1 | 5,997 | 24.0 |
| 582 | 526 | 2.1 | 6,523 | 26.1 |
| 583 | 535 | 2.1 | 7,058 | 28.3 |
| 584 | 378 | 1.5 | 7,436 | 29.8 |
| 585 | 493 | 2.0 | 7,929 | 31.8 |
| 586 | 359 | 1.4 | 8,288 | 33.2 |
| 587 | 736 | 2.9 | 9,024 | 36.1 |
| 588 | 224 | 0.9 | 9,248 | 37.0 |
| 589 | 821 | 3.3 | 10,069 | 40.3 |
| 590 | 140 | 0.6 | 10,209 | 40.9 |
| 591 | 761 | 3.0 | 10,970 | 43.9 |
| 592 | 108 | 0.4 | 11,078 | 44.4 |
| 593 | 696 | 2.8 | 11,774 | 47.1 |
| 594 | 97 | 0.4 | 11,871 | 47.5 |
| 595 | 708 | 2.8 | 12,579 | 50.4 |
| 596 | 92 | 0.4 | 12,671 | 50.7 |
| 597 | 663 | 2.7 | 13,334 | 53.4 |
| 598 | 100 | 0.4 | 13,434 | 53.8 |
| 599 | 625 | 2.5 | 14,059 | 56.3 |
| 600 | 284 | 1.1 | 14,343 | 57.4 |
| 601 | 389 | 1.6 | 14,732 | 59.0 |
| 602 | 360 | 1.4 | 15,092 | 60.4 |
| 603 | 370 | 1.5 | 15,462 | 61.9 |
| 604 | 387 | 1.5 | 15,849 | 63.5 |
| 605 | 248 | 1.0 | 16,097 | 64.5 |
| 606 | 381 | 1.5 | 16,478 | 66.0 |
| 607 | 237 | 0.9 | 16,715 | 66.9 |
| 608 | 373 | 1.5 | 17,088 | 68.4 |
| 609 | 251 | 1.0 | 17,339 | 69.4 |
| 610 | 359 | 1.4 | 17,698 | 70.9 |
| 611 | 309 | 1.2 | 18,007 | 72.1 |
| 612 | 254 | 1.0 | 18,261 | 73.1 |
| 613 | 372 | 1.5 | 18,633 | 74.6 |
| 614 | 197 | 0.8 | 18,830 | 75.4 |
| 615 | 506 | 2.0 | 19,336 | 77.4 |
| 616 | 2 | <0.1 | 19,338 | 77.4 |
| 617 | 531 | 2.1 | 19,869 | 79.6 |
| 618 | 0 | 0.0 | 19,869 | 79.6 |
| 619 | 487 | 2.0 | 20,356 | 81.5 |
| 620 | 109 | 0.4 | 20,465 | 82.0 |
| 621 | 340 | 1.4 | 20,805 | 83.3 |
| 622 | 325 | 1.3 | 21,130 | 84.6 |
| 623 | 105 | 0.4 | 21,235 | 85.0 |
| 624 | 434 | 1.7 | 21,669 | 86.8 |
| 625 | 1 | <0.1 | 21,670 | 86.8 |
| 626 | 416 | 1.7 | 22,086 | 88.4 |
| 627 | 175 | 0.7 | 22,261 | 89.1 |
| 628 | 201 | 0.8 | 22,462 | 89.9 |
| 629 | 229 | 0.9 | 22,691 | 90.9 |
| 630 | 127 | 0.5 | 22,818 | 91.4 |
| 631 | 271 | 1.1 | 23,089 | 92.5 |
| 632 | 150 | 0.6 | 23,239 | 93.1 |
| 633 | 152 | 0.6 | 23,391 | 93.7 |
| 634 | 220 | 0.9 | 23,611 | 94.5 |
| 635 | 133 | 0.5 | 23,744 | 95.1 |
| 636 | 152 | 0.6 | 23,896 | 95.7 |
| 637 | 158 | 0.6 | 24,054 | 96.3 |
| 638 | 97 | 0.4 | 24,151 | 96.7 |
| 639 | 133 | 0.5 | 24,284 | 97.2 |
| 640 | 117 | 0.5 | 24,401 | 97.7 |
| 641 | 78 | 0.3 | 24,479 | 98.0 |
| 642 | 73 | 0.3 | 24,552 | 98.3 |
| 643 | 115 | 0.5 | 24,667 | 98.8 |
| 644 | 54 | 0.2 | 24,721 | 99.0 |
| 645 | 82 | 0.3 | 24,803 | 99.3 |
| 646 | 38 | 0.2 | 24,841 | 99.5 |
| 647 | 56 | 0.2 | 24,897 | 99.7 |
| 648 | 21 | <0.1 | 24,918 | 99.8 |
| 649 | 26 | 0.1 | 24,944 | 99.9 |
| 650 | 28 | 0.1 | 24,972 | 100.0 |

Table 7.B.4 CAST Grade Eleven Scale Scores

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scale Score** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 550 | 468 | 0.1 | 468 | 0.1 |
| 551 | 4 | <0.1 | 472 | 0.1 |
| 552 | 10 | <0.1 | 482 | 0.1 |
| 553 | 22 | <0.1 | 504 | 0.2 |
| 554 | 25 | <0.1 | 529 | 0.2 |
| 555 | 21 | <0.1 | 550 | 0.2 |
| 556 | 116 | <0.1 | 666 | 0.2 |
| 557 | 45 | <0.1 | 711 | 0.2 |
| 558 | 71 | <0.1 | 782 | 0.2 |
| 559 | 344 | 0.1 | 1,126 | 0.3 |
| 560 | 58 | <0.1 | 1,184 | 0.4 |
| 561 | 832 | 0.3 | 2,016 | 0.6 |
| 562 | 75 | <0.1 | 2,091 | 0.6 |
| 563 | 824 | 0.2 | 2,915 | 0.9 |
| 564 | 1,011 | 0.3 | 3,926 | 1.2 |
| 565 | 436 | 0.1 | 4,362 | 1.3 |
| 566 | 2,938 | 0.9 | 7,300 | 2.2 |
| 567 | 58 | <0.1 | 7,358 | 2.2 |
| 568 | 3,121 | 0.9 | 10,479 | 3.2 |
| 569 | 2,134 | 0.6 | 12,613 | 3.8 |
| 570 | 3,381 | 1.0 | 15,994 | 4.8 |
| 571 | 3,910 | 1.2 | 19,904 | 6.0 |
| 572 | 3,220 | 1.0 | 23,124 | 7.0 |
| 573 | 5,957 | 1.8 | 29,081 | 8.7 |
| 574 | 3,669 | 1.1 | 32,750 | 9.9 |
| 575 | 3,939 | 1.2 | 36,689 | 11.0 |
| 576 | 5,545 | 1.7 | 42,234 | 12.7 |
| 577 | 5,731 | 1.7 | 47,965 | 14.4 |
| 578 | 5,786 | 1.7 | 53,751 | 16.2 |
| 579 | 5,925 | 1.8 | 59,676 | 18.0 |
| 580 | 5,945 | 1.8 | 65,621 | 19.7 |
| 581 | 6,054 | 1.8 | 71,675 | 21.6 |
| 582 | 5,907 | 1.8 | 77,582 | 23.3 |
| 583 | 5,778 | 1.7 | 83,360 | 25.1 |
| 584 | 4,174 | 1.3 | 87,534 | 26.3 |
| 585 | 5,828 | 1.8 | 93,362 | 28.1 |
| 586 | 3,761 | 1.1 | 97,123 | 29.2 |
| 587 | 8,370 | 2.5 | 105,493 | 31.7 |
| 588 | 2,559 | 0.8 | 108,052 | 32.5 |
| 589 | 9,117 | 2.7 | 117,169 | 35.2 |
| 590 | 1,306 | 0.4 | 118,475 | 35.6 |
| 591 | 8,785 | 2.6 | 127,260 | 38.3 |
| 592 | 1,223 | 0.4 | 128,483 | 38.7 |
| 593 | 8,385 | 2.5 | 136,868 | 41.2 |
| 594 | 1,190 | 0.4 | 138,058 | 41.5 |
| 595 | 8,090 | 2.4 | 146,148 | 44.0 |
| 596 | 1,116 | 0.3 | 147,264 | 44.3 |
| 597 | 7,956 | 2.4 | 155,220 | 46.7 |
| 598 | 1,038 | 0.3 | 156,258 | 47.0 |
| 599 | 7,559 | 2.3 | 163,817 | 49.3 |
| 600 | 3,329 | 1.0 | 167,146 | 50.3 |
| 601 | 5,378 | 1.6 | 172,524 | 51.9 |
| 602 | 4,225 | 1.3 | 176,749 | 53.2 |
| 603 | 4,218 | 1.3 | 180,967 | 54.4 |
| 604 | 5,078 | 1.5 | 186,045 | 56.0 |
| 605 | 3,153 | 0.9 | 189,198 | 56.9 |
| 606 | 4,945 | 1.5 | 194,143 | 58.4 |
| 607 | 3,222 | 1.0 | 197,365 | 59.4 |
| 608 | 4,978 | 1.5 | 202,343 | 60.9 |
| 609 | 3,135 | 0.9 | 205,478 | 61.8 |
| 610 | 4,951 | 1.5 | 210,429 | 63.3 |
| 611 | 3,962 | 1.2 | 214,391 | 64.5 |
| 612 | 3,887 | 1.2 | 218,278 | 65.7 |
| 613 | 4,813 | 1.4 | 223,091 | 67.1 |
| 614 | 2,821 | 0.8 | 225,912 | 68.0 |
| 615 | 7,479 | 2.2 | 233,391 | 70.2 |
| 616 | 10 | <0.1 | 233,401 | 70.2 |
| 617 | 7,370 | 2.2 | 240,771 | 72.4 |
| 618 | 12 | <0.1 | 240,783 | 72.4 |
| 619 | 7,288 | 2.2 | 248,071 | 74.6 |
| 620 | 1,852 | 0.6 | 249,923 | 75.2 |
| 621 | 5,400 | 1.6 | 255,323 | 76.8 |
| 622 | 5,241 | 1.6 | 260,564 | 78.4 |
| 623 | 1,771 | 0.5 | 262,335 | 78.9 |
| 624 | 6,828 | 2.1 | 269,163 | 81.0 |
| 625 | 5 | <0.1 | 269,168 | 81.0 |
| 626 | 6,464 | 1.9 | 275,632 | 82.9 |
| 627 | 3,039 | 0.9 | 278,671 | 83.8 |
| 628 | 3,257 | 1.0 | 281,928 | 84.8 |
| 629 | 4,520 | 1.4 | 286,448 | 86.2 |
| 630 | 2,234 | 0.7 | 288,682 | 86.8 |
| 631 | 5,088 | 1.5 | 293,770 | 88.4 |
| 632 | 2,636 | 0.8 | 296,406 | 89.2 |
| 633 | 2,651 | 0.8 | 299,057 | 90.0 |
| 634 | 4,285 | 1.3 | 303,342 | 91.3 |
| 635 | 2,377 | 0.7 | 305,719 | 92.0 |
| 636 | 2,876 | 0.9 | 308,595 | 92.8 |
| 637 | 3,193 | 1.0 | 311,788 | 93.8 |
| 638 | 2,017 | 0.6 | 313,805 | 94.4 |
| 639 | 2,901 | 0.9 | 316,706 | 95.3 |
| 640 | 2,992 | 0.9 | 319,698 | 96.2 |
| 641 | 1,903 | 0.6 | 321,601 | 96.7 |
| 642 | 1,746 | 0.5 | 323,347 | 97.3 |
| 643 | 2,065 | 0.6 | 325,412 | 97.9 |
| 644 | 1,068 | 0.3 | 326,480 | 98.2 |
| 645 | 1,827 | 0.5 | 328,307 | 98.8 |
| 646 | 1,119 | 0.3 | 329,426 | 99.1 |
| 647 | 1,194 | 0.4 | 330,620 | 99.5 |
| 648 | 550 | 0.2 | 331,170 | 99.6 |
| 649 | 613 | 0.2 | 331,783 | 99.8 |
| 650 | 634 | 0.2 | 332,417 | 100.0 |

Table 7.B.5 CAST Grade Twelve Scale Scores

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scale Score** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 550 | 203 | 0.2 | 203 | 0.2 |
| 551 | 1 | <0.1 | 204 | 0.2 |
| 552 | 6 | <0.1 | 210 | 0.2 |
| 553 | 9 | <0.1 | 219 | 0.2 |
| 554 | 19 | <0.1 | 238 | 0.2 |
| 555 | 10 | <0.1 | 248 | 0.2 |
| 556 | 29 | <0.1 | 277 | 0.3 |
| 557 | 21 | <0.1 | 298 | 0.3 |
| 558 | 40 | <0.1 | 338 | 0.3 |
| 559 | 150 | 0.2 | 488 | 0.5 |
| 560 | 16 | <0.1 | 504 | 0.5 |
| 561 | 385 | 0.4 | 889 | 0.9 |
| 562 | 25 | <0.1 | 914 | 0.9 |
| 563 | 353 | 0.4 | 1,267 | 1.3 |
| 564 | 460 | 0.5 | 1,727 | 1.7 |
| 565 | 171 | 0.2 | 1,898 | 1.9 |
| 566 | 1,258 | 1.3 | 3,156 | 3.2 |
| 567 | 23 | <0.1 | 3,179 | 3.2 |
| 568 | 1,280 | 1.3 | 4,459 | 4.5 |
| 569 | 895 | 0.9 | 5,354 | 5.4 |
| 570 | 1,370 | 1.4 | 6,724 | 6.7 |
| 571 | 1,567 | 1.6 | 8,291 | 8.3 |
| 572 | 1,233 | 1.2 | 9,524 | 9.5 |
| 573 | 2,127 | 2.1 | 11,651 | 11.7 |
| 574 | 1,402 | 1.4 | 13,053 | 13.1 |
| 575 | 1,439 | 1.4 | 14,492 | 14.5 |
| 576 | 2,019 | 2.0 | 16,511 | 16.5 |
| 577 | 2,087 | 2.1 | 18,598 | 18.6 |
| 578 | 2,170 | 2.2 | 20,768 | 20.8 |
| 579 | 2,188 | 2.2 | 22,956 | 23.0 |
| 580 | 2,074 | 2.1 | 25,030 | 25.1 |
| 581 | 2,059 | 2.1 | 27,089 | 27.1 |
| 582 | 2,058 | 2.1 | 29,147 | 29.2 |
| 583 | 1,987 | 2.0 | 31,134 | 31.2 |
| 584 | 1,389 | 1.4 | 32,523 | 32.6 |
| 585 | 1,945 | 1.9 | 34,468 | 34.5 |
| 586 | 1,260 | 1.3 | 35,728 | 35.8 |
| 587 | 2,733 | 2.7 | 38,461 | 38.5 |
| 588 | 839 | 0.8 | 39,300 | 39.4 |
| 589 | 2,889 | 2.9 | 42,189 | 42.3 |
| 590 | 400 | 0.4 | 42,589 | 42.7 |
| 591 | 2,734 | 2.7 | 45,323 | 45.4 |
| 592 | 382 | 0.4 | 45,705 | 45.8 |
| 593 | 2,587 | 2.6 | 48,292 | 48.4 |
| 594 | 351 | 0.4 | 48,643 | 48.7 |
| 595 | 2,530 | 2.5 | 51,173 | 51.3 |
| 596 | 362 | 0.4 | 51,535 | 51.6 |
| 597 | 2,355 | 2.4 | 53,890 | 54.0 |
| 598 | 325 | 0.3 | 54,215 | 54.3 |
| 599 | 2,213 | 2.2 | 56,428 | 56.5 |
| 600 | 947 | 0.9 | 57,375 | 57.5 |
| 601 | 1,524 | 1.5 | 58,899 | 59.0 |
| 602 | 1,202 | 1.2 | 60,101 | 60.2 |
| 603 | 1,158 | 1.2 | 61,259 | 61.4 |
| 604 | 1,396 | 1.4 | 62,655 | 62.8 |
| 605 | 908 | 0.9 | 63,563 | 63.7 |
| 606 | 1,415 | 1.4 | 64,978 | 65.1 |
| 607 | 896 | 0.9 | 65,874 | 66.0 |
| 608 | 1,346 | 1.3 | 67,220 | 67.4 |
| 609 | 797 | 0.8 | 68,017 | 68.1 |
| 610 | 1,243 | 1.2 | 69,260 | 69.4 |
| 611 | 1,029 | 1.0 | 70,289 | 70.4 |
| 612 | 941 | 0.9 | 71,230 | 71.4 |
| 613 | 1,251 | 1.3 | 72,481 | 72.6 |
| 614 | 701 | 0.7 | 73,182 | 73.3 |
| 615 | 1,922 | 1.9 | 75,104 | 75.3 |
| 616 | 5 | <0.1 | 75,109 | 75.3 |
| 617 | 1,901 | 1.9 | 77,010 | 77.2 |
| 618 | 5 | <0.1 | 77,015 | 77.2 |
| 619 | 1,770 | 1.8 | 78,785 | 78.9 |
| 620 | 450 | 0.5 | 79,235 | 79.4 |
| 621 | 1,318 | 1.3 | 80,553 | 80.7 |
| 622 | 1,269 | 1.3 | 81,822 | 82.0 |
| 623 | 413 | 0.4 | 82,235 | 82.4 |
| 624 | 1,662 | 1.7 | 83,897 | 84.1 |
| 625 | 3 | <0.1 | 83,900 | 84.1 |
| 626 | 1,591 | 1.6 | 85,491 | 85.7 |
| 627 | 745 | 0.7 | 86,236 | 86.4 |
| 628 | 774 | 0.8 | 87,010 | 87.2 |
| 629 | 1,150 | 1.2 | 88,160 | 88.3 |
| 630 | 529 | 0.5 | 88,689 | 88.9 |
| 631 | 1,286 | 1.3 | 89,975 | 90.2 |
| 632 | 624 | 0.6 | 90,599 | 90.8 |
| 633 | 627 | 0.6 | 91,226 | 91.4 |
| 634 | 1,021 | 1.0 | 92,247 | 92.4 |
| 635 | 563 | 0.6 | 92,810 | 93.0 |
| 636 | 704 | 0.7 | 93,514 | 93.7 |
| 637 | 775 | 0.8 | 94,289 | 94.5 |
| 638 | 509 | 0.5 | 94,798 | 95.0 |
| 639 | 705 | 0.7 | 95,503 | 95.7 |
| 640 | 775 | 0.8 | 96,278 | 96.5 |
| 641 | 495 | 0.5 | 96,773 | 97.0 |
| 642 | 477 | 0.5 | 97,250 | 97.4 |
| 643 | 594 | 0.6 | 97,844 | 98.0 |
| 644 | 262 | 0.3 | 98,106 | 98.3 |
| 645 | 532 | 0.5 | 98,638 | 98.8 |
| 646 | 287 | 0.3 | 98,925 | 99.1 |
| 647 | 350 | 0.4 | 99,275 | 99.5 |
| 648 | 158 | 0.2 | 99,433 | 99.6 |
| 649 | 183 | 0.2 | 99,616 | 99.8 |
| 650 | 189 | 0.2 | 99,805 | 100.0 |

Table 7.B.6 CAST High School Scale Scores

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scale Score** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 550 | 708 | 0.2 | 708 | 0.2 |
| 551 | 5 | <0.1 | 713 | 0.2 |
| 552 | 16 | <0.1 | 729 | 0.2 |
| 553 | 32 | <0.1 | 761 | 0.2 |
| 554 | 47 | <0.1 | 808 | 0.2 |
| 555 | 33 | <0.1 | 841 | 0.2 |
| 556 | 159 | <0.1 | 1,000 | 0.2 |
| 557 | 69 | <0.1 | 1,069 | 0.2 |
| 558 | 122 | <0.1 | 1,191 | 0.3 |
| 559 | 516 | 0.1 | 1,707 | 0.4 |
| 560 | 79 | <0.1 | 1,786 | 0.4 |
| 561 | 1,276 | 0.3 | 3,062 | 0.7 |
| 562 | 108 | <0.1 | 3,170 | 0.7 |
| 563 | 1,250 | 0.3 | 4,420 | 1.0 |
| 564 | 1,570 | 0.3 | 5,990 | 1.3 |
| 565 | 635 | 0.1 | 6,625 | 1.4 |
| 566 | 4,423 | 1.0 | 11,048 | 2.4 |
| 567 | 83 | <0.1 | 11,131 | 2.4 |
| 568 | 4,613 | 1.0 | 15,744 | 3.4 |
| 569 | 3,184 | 0.7 | 18,928 | 4.1 |
| 570 | 5,036 | 1.1 | 23,964 | 5.2 |
| 571 | 5,798 | 1.3 | 29,762 | 6.5 |
| 572 | 4,755 | 1.0 | 34,517 | 7.5 |
| 573 | 8,533 | 1.9 | 43,050 | 9.4 |
| 574 | 5,401 | 1.2 | 48,451 | 10.6 |
| 575 | 5,691 | 1.2 | 54,142 | 11.8 |
| 576 | 8,046 | 1.8 | 62,188 | 13.6 |
| 577 | 8,291 | 1.8 | 70,479 | 15.4 |
| 578 | 8,484 | 1.9 | 78,963 | 17.3 |
| 579 | 8,625 | 1.9 | 87,588 | 19.2 |
| 580 | 8,539 | 1.9 | 96,127 | 21.0 |
| 581 | 8,634 | 1.9 | 104,761 | 22.9 |
| 582 | 8,491 | 1.9 | 113,252 | 24.8 |
| 583 | 8,300 | 1.8 | 121,552 | 26.6 |
| 584 | 5,941 | 1.3 | 127,493 | 27.9 |
| 585 | 8,266 | 1.8 | 135,759 | 29.7 |
| 586 | 5,380 | 1.2 | 141,139 | 30.9 |
| 587 | 11,839 | 2.6 | 152,978 | 33.5 |
| 588 | 3,622 | 0.8 | 156,600 | 34.3 |
| 589 | 12,827 | 2.8 | 169,427 | 37.1 |
| 590 | 1,846 | 0.4 | 171,273 | 37.5 |
| 591 | 12,280 | 2.7 | 183,553 | 40.1 |
| 592 | 1,713 | 0.4 | 185,266 | 40.5 |
| 593 | 11,668 | 2.6 | 196,934 | 43.1 |
| 594 | 1,638 | 0.4 | 198,572 | 43.4 |
| 595 | 11,328 | 2.5 | 209,900 | 45.9 |
| 596 | 1,570 | 0.3 | 211,470 | 46.3 |
| 597 | 10,974 | 2.4 | 222,444 | 48.7 |
| 598 | 1,463 | 0.3 | 223,907 | 49.0 |
| 599 | 10,397 | 2.3 | 234,304 | 51.2 |
| 600 | 4,560 | 1.0 | 238,864 | 52.2 |
| 601 | 7,291 | 1.6 | 246,155 | 53.8 |
| 602 | 5,787 | 1.3 | 251,942 | 55.1 |
| 603 | 5,746 | 1.3 | 257,688 | 56.4 |
| 604 | 6,861 | 1.5 | 264,549 | 57.9 |
| 605 | 4,309 | 0.9 | 268,858 | 58.8 |
| 606 | 6,741 | 1.5 | 275,599 | 60.3 |
| 607 | 4,355 | 1.0 | 279,954 | 61.2 |
| 608 | 6,697 | 1.5 | 286,651 | 62.7 |
| 609 | 4,183 | 0.9 | 290,834 | 63.6 |
| 610 | 6,553 | 1.4 | 297,387 | 65.0 |
| 611 | 5,300 | 1.2 | 302,687 | 66.2 |
| 612 | 5,082 | 1.1 | 307,769 | 67.3 |
| 613 | 6,436 | 1.4 | 314,205 | 68.7 |
| 614 | 3,719 | 0.8 | 317,924 | 69.5 |
| 615 | 9,907 | 2.2 | 327,831 | 71.7 |
| 616 | 17 | <0.1 | 327,848 | 71.7 |
| 617 | 9,802 | 2.1 | 337,650 | 73.9 |
| 618 | 17 | <0.1 | 337,667 | 73.9 |
| 619 | 9,545 | 2.1 | 347,212 | 75.9 |
| 620 | 2,411 | 0.5 | 349,623 | 76.5 |
| 621 | 7,058 | 1.5 | 356,681 | 78.0 |
| 622 | 6,835 | 1.5 | 363,516 | 79.5 |
| 623 | 2,289 | 0.5 | 365,805 | 80.0 |
| 624 | 8,924 | 2.0 | 374,729 | 82.0 |
| 625 | 9 | <0.1 | 374,738 | 82.0 |
| 626 | 8,471 | 1.9 | 383,209 | 83.8 |
| 627 | 3,959 | 0.9 | 387,168 | 84.7 |
| 628 | 4,232 | 0.9 | 391,400 | 85.6 |
| 629 | 5,899 | 1.3 | 397,299 | 86.9 |
| 630 | 2,890 | 0.6 | 400,189 | 87.5 |
| 631 | 6,645 | 1.5 | 406,834 | 89.0 |
| 632 | 3,410 | 0.7 | 410,244 | 89.7 |
| 633 | 3,430 | 0.8 | 413,674 | 90.5 |
| 634 | 5,526 | 1.2 | 419,200 | 91.7 |
| 635 | 3,073 | 0.7 | 422,273 | 92.4 |
| 636 | 3,732 | 0.8 | 426,005 | 93.2 |
| 637 | 4,126 | 0.9 | 430,131 | 94.1 |
| 638 | 2,623 | 0.6 | 432,754 | 94.7 |
| 639 | 3,739 | 0.8 | 436,493 | 95.5 |
| 640 | 3,884 | 0.8 | 440,377 | 96.3 |
| 641 | 2,476 | 0.5 | 442,853 | 96.9 |
| 642 | 2,296 | 0.5 | 445,149 | 97.4 |
| 643 | 2,774 | 0.6 | 447,923 | 98.0 |
| 644 | 1,384 | 0.3 | 449,307 | 98.3 |
| 645 | 2,441 | 0.5 | 451,748 | 98.8 |
| 646 | 1,444 | 0.3 | 453,192 | 99.1 |
| 647 | 1,600 | 0.3 | 454,792 | 99.5 |
| 648 | 729 | 0.2 | 455,521 | 99.6 |
| 649 | 822 | 0.2 | 456,343 | 99.8 |
| 650 | 851 | 0.2 | 457,194 | 100.0 |

### Appendix 7.C: Demographic Student Group Summaries of Participation

**Notes:**

* The following tables are based on students with valid test scores and are presented separately for grades five and eight and for high school, which shows categories for grades ten, eleven, and twelve.
* All students are required to take CAST in grades five and eight. High school students are required to test once in grade ten, eleven, or twelve. Because CAST is a census test for students in grades five and eight, but not a census test for high school students, two student data files were used to calculate the percentages in table 7.C.1 through table 7.C.6: the California Longitudinal Pupil Achievement Data System (CALPADS) and TOMS. CALPADS was used for grades five and eight, while CALPADS and the registered student information in TOMS were used for high school. The TOMS file was required for high school to be able to determine whether students were expected to test in this administration.
* The *Percent of Valid Scores* calculations are based on test results. The *Population Percent* calculations are based on both CALPADS and, for high school, TOMS registration data.
* The percentages of student groups may not sum to 100 because of rounding. In addition, the percentages of students within a race or ethnicity category may not sum to 100 because of missing race or ethnicity data for some students.

Table 7.C.1 Demographic Student Group Summaries for Grade Five

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** | **Population Percent** |
| All students | 422,398 | 100.0 | 100.0 |
| Male | 215,552 | 51.0 | 51.3 |
| Female | 206,730 | 48.9 | 48.7 |
| Nonbinary | 116 | <0.1 | <0.1 |
| EL | 80,657 | 19.1 | 19.6 |
| English only | 259,625 | 61.5 | 61.7 |
| RFEP | 55,501 | 13.1 | 12.5 |
| IFEP | 26,483 | 6.3 | 6.1 |
| ADEL | 0 | 0.0 | 0.0 |
| To be determined | 69 | <0.1 | <0.1 |
| English proficiency unknown | 63 | <0.1 | <0.1 |
| Economically disadvantaged | 277,306 | 65.7 | 64.4 |
| Not economically disadvantaged | 145,092 | 34.3 | 35.6 |
| American Indian or Alaska Native (All) | 1,807 | 0.4 | 0.4 |
| Asian (All) | 44,954 | 10.6 | 10.5 |
| Native Hawaiian or Other Pacific Islander (All) | 1,708 | 0.4 | 0.4 |
| Filipino (All) | 9,411 | 2.2 | 2.2 |
| Hispanic or Latino (All) | 234,953 | 55.6 | 55.4 |
| Black or African American (All) | 20,881 | 4.9 | 5.0 |
| White (All) | 84,422 | 20.0 | 20.1 |
| Two or more races (All) | 24,262 | 5.7 | 4.7 |
| Disability | 56,571 | 13.4 | 15.0 |
| No disability | 365,827 | 86.6 | 85.0 |
| Migrant education | 3,438 | 0.8 | 0.8 |
| Not migrant education | 418,960 | 99.2 | 99.2 |
| Armed forces family member | 6,493 | 1.5 | 1.5 |
| Not armed forces family member | 415,905 | 98.5 | 98.5 |
| Homeless | 15,571 | 3.7 | 4.3 |
| Not homeless | 406,827 | 96.3 | 95.7 |
| Foster youth | 1,355 | 0.3 | 0.3 |
| Not foster youth | 421,043 | 99.7 | 99.7 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 466 | 0.1 | 0.1 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 1,341 | 0.3 | 0.3 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 25,678 | 6.1 | 6.2 |
| Asian (Primary ethnicity—Economically disadvantaged) | 19,276 | 4.6 | 4.3 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 425 | 0.1 | 0.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 1,283 | 0.3 | 0.3 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 5,313 | 1.3 | 1.3 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 4,098 | 1.0 | 0.9 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 45,455 | 10.8 | 11.0 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 189,498 | 44.9 | 44.4 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 4,138 | 1.0 | 1.0 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 16,743 | 4.0 | 4.0 |
| White (Primary ethnicity—Not economically disadvantaged) | 50,304 | 11.9 | 12.4 |
| White (Primary ethnicity—Economically disadvantaged) | 34,118 | 8.1 | 7.6 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 13,313 | 3.2 | 2.8 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 10,949 | 2.6 | 1.9 |

Table 7.C.2 Demographic Student Group Summaries for Grade Eight

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** | **Population Percent** |
| All students | 427,448 | 100.0 | 100.0 |
| Male | 219,022 | 51.2 | 51.4 |
| Female | 208,044 | 48.7 | 48.5 |
| Nonbinary | 382 | <0.1 | <0.1 |
| EL | 54,363 | 12.7 | 13.1 |
| English only | 243,184 | 56.9 | 57.4 |
| RFEP | 111,588 | 26.1 | 25.2 |
| IFEP | 18,220 | 4.3 | 4.2 |
| ADEL | 0 | 0.0 | 0.0 |
| To be determined | 57 | <0.1 | <0.1 |
| English proficiency unknown | 36 | <0.1 | <0.1 |
| Economically disadvantaged | 278,202 | 65.1 | 64.2 |
| Not economically disadvantaged | 149,246 | 34.9 | 35.8 |
| American Indian or Alaska Native (All) | 1,806 | 0.4 | 0.4 |
| Asian (All) | 42,481 | 9.9 | 9.7 |
| Native Hawaiian or Other Pacific Islander (All) | 1,852 | 0.4 | 0.4 |
| Filipino (All) | 10,074 | 2.4 | 2.3 |
| Hispanic or Latino (All) | 242,458 | 56.7 | 56.5 |
| Black or African American (All) | 21,441 | 5.0 | 5.1 |
| White (All) | 85,485 | 20.0 | 20.2 |
| Two or more races (All) | 21,851 | 5.1 | 4.3 |
| Disability | 50,652 | 11.8 | 13.5 |
| No disability | 376,796 | 88.2 | 86.5 |
| Migrant education | 3,341 | 0.8 | 0.8 |
| Not migrant education | 424,107 | 99.2 | 99.2 |
| Armed forces family member | 5,989 | 1.4 | 1.3 |
| Not armed forces family member | 421,459 | 98.6 | 98.7 |
| Homeless | 13,775 | 3.2 | 3.8 |
| Not homeless | 413,673 | 96.8 | 96.2 |
| Foster youth | 1,431 | 0.3 | 0.4 |
| Not foster youth | 426,017 | 99.7 | 99.6 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 527 | 0.1 | 0.1 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 1,279 | 0.3 | 0.3 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 24,606 | 5.8 | 5.7 |
| Asian (Primary ethnicity—Economically disadvantaged) | 17,875 | 4.2 | 4.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 518 | 0.1 | 0.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 1,334 | 0.3 | 0.3 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 5,934 | 1.4 | 1.4 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 4,140 | 1.0 | 0.9 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 48,524 | 11.4 | 11.6 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 193,934 | 45.4 | 44.9 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 4,894 | 1.1 | 1.2 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 16,547 | 3.9 | 3.9 |
| White (Primary ethnicity—Not economically disadvantaged) | 51,833 | 12.1 | 12.6 |
| White (Primary ethnicity—Economically disadvantaged) | 33,652 | 7.9 | 7.6 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 12,410 | 2.9 | 2.6 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 9,441 | 2.2 | 1.7 |

Table 7.C.3 Demographic Student Group Summaries for Grade Ten

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** | **Population Percent** |
| All students | 24,935 | 100.0 | 100.0 |
| Male | 12,811 | 51.4 | 51.3 |
| Female | 12,060 | 48.4 | 48.1 |
| Nonbinary | 64 | 0.3 | 0.3 |
| EL | 3,070 | 12.3 | 12.7 |
| English only | 13,756 | 55.2 | 55.2 |
| RFEP | 7,077 | 28.4 | 27.9 |
| IFEP | 986 | 4.0 | 3.8 |
| ADEL | 41 | 0.2 | <0.1 |
| To be determined | 0 | 0.0 | <0.1 |
| English proficiency unknown | 5 | <0.1 | 0.4 |
| Economically disadvantaged | 16,713 | 67.0 | 67.2 |
| Not economically disadvantaged | 8,222 | 33.0 | 32.5 |
| American Indian or Alaska Native (All) | 129 | 0.5 | 0.5 |
| Asian (All) | 1,638 | 6.6 | 6.0 |
| Native Hawaiian or Other Pacific Islander (All) | 84 | 0.3 | 0.3 |
| Filipino (All) | 560 | 2.2 | 2.0 |
| Hispanic or Latino (All) | 15,379 | 61.7 | 62.9 |
| Black or African American (All) | 1,463 | 5.9 | 5.4 |
| White (All) | 4,674 | 18.7 | 18.1 |
| Two or more races (All) | 1,008 | 4.0 | 3.6 |
| Disability | 3,090 | 12.4 | 13.3 |
| No disability | 21,845 | 87.6 | 86.5 |
| Migrant education | 233 | 0.9 | 0.9 |
| Not migrant education | 24,702 | 99.1 | 98.8 |
| Armed forces family member | 314 | 1.3 | 1.4 |
| Not armed forces family member | 24,621 | 98.7 | 98.3 |
| Homeless | 758 | 3.0 | 3.2 |
| Not homeless | 24,177 | 97.0 | 96.5 |
| Foster youth | 150 | 0.6 | 0.6 |
| Not foster youth | 24,785 | 99.4 | 99.4 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 39 | 0.2 | 0.1 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 90 | 0.4 | 0.3 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 889 | 3.6 | 3.3 |
| Asian (Primary ethnicity—Economically disadvantaged) | 749 | 3.0 | 2.8 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 35 | 0.1 | 0.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 49 | 0.2 | 0.2 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 352 | 1.4 | 1.2 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 208 | 0.8 | 0.7 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 3,221 | 12.9 | 13.2 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 12,158 | 48.8 | 49.7 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 364 | 1.5 | 1.4 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 1,099 | 4.4 | 4.0 |
| White (Primary ethnicity—Not economically disadvantaged) | 2,761 | 11.1 | 10.7 |
| White (Primary ethnicity—Economically disadvantaged) | 1,913 | 7.7 | 7.4 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 561 | 2.2 | 2.0 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 447 | 1.8 | 1.5 |

Table 7.C.4 Demographic Student Group Summaries for Grade Eleven

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** | **Population Percent** |
| All students | 331,949 | 100.0 | 100.0 |
| Male | 169,911 | 51.2 | 51.0 |
| Female | 161,365 | 48.6 | 48.5 |
| Nonbinary | 673 | 0.2 | 0.2 |
| EL | 35,104 | 10.6 | 11.0 |
| English only | 178,868 | 53.9 | 54.2 |
| RFEP | 102,236 | 30.8 | 29.9 |
| IFEP | 15,690 | 4.7 | 4.6 |
| ADEL | 0 | 0.0 | 0.0 |
| To be determined | 26 | <0.1 | <0.1 |
| English proficiency unknown | 25 | <0.1 | 0.3 |
| Economically disadvantaged | 206,636 | 62.2 | 62.3 |
| Not economically disadvantaged | 125,313 | 37.8 | 37.5 |
| American Indian or Alaska Native (All) | 1,427 | 0.4 | 0.4 |
| Asian (All) | 33,303 | 10.0 | 9.8 |
| Native Hawaiian or Other Pacific Islander (All) | 1,411 | 0.4 | 0.4 |
| Filipino (All) | 8,625 | 2.6 | 2.5 |
| Hispanic or Latino (All) | 187,695 | 56.5 | 56.1 |
| Black or African American (All) | 14,690 | 4.4 | 4.6 |
| White (All) | 69,609 | 21.0 | 21.2 |
| Two or more races (All) | 15,189 | 4.6 | 3.9 |
| Disability | 35,597 | 10.7 | 11.5 |
| No disability | 296,352 | 89.3 | 88.3 |
| Migrant education | 2,252 | 0.7 | 0.7 |
| Not migrant education | 329,697 | 99.3 | 99.1 |
| Armed forces family member | 4,971 | 1.5 | 1.4 |
| Not armed forces family member | 326,978 | 98.5 | 98.3 |
| Homeless | 10,109 | 3.0 | 3.4 |
| Not homeless | 321,840 | 97.0 | 96.3 |
| Foster youth | 1,133 | 0.3 | 0.4 |
| Not foster youth | 330,816 | 99.7 | 99.6 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 503 | 0.2 | 0.1 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 924 | 0.3 | 0.3 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 18,754 | 5.6 | 5.5 |
| Asian (Primary ethnicity—Economically disadvantaged) | 14,549 | 4.4 | 4.3 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 508 | 0.2 | 0.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 903 | 0.3 | 0.3 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 5,182 | 1.6 | 1.5 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 3,443 | 1.0 | 1.0 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 41,355 | 12.5 | 12.3 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 146,340 | 44.1 | 43.8 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 4,194 | 1.3 | 1.3 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 10,496 | 3.2 | 3.3 |
| White (Primary ethnicity—Not economically disadvantaged) | 45,626 | 13.7 | 13.8 |
| White (Primary ethnicity—Economically disadvantaged) | 23,983 | 7.2 | 7.3 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 9,191 | 2.8 | 2.4 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 5,998 | 1.8 | 1.5 |

Table 7.C.5 Demographic Student Group Summaries for Grade Twelve

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** | **Population Percent** |
| All students | 99,602 | 100.0 | 100.0 |
| Male | 50,938 | 51.1 | 49.5 |
| Female | 48,452 | 48.6 | 47.5 |
| Nonbinary | 212 | 0.2 | 0.2 |
| EL | 12,020 | 12.1 | 12.4 |
| English only | 50,890 | 51.1 | 51.1 |
| RFEP | 31,068 | 31.2 | 27.9 |
| IFEP | 5,377 | 5.4 | 5.0 |
| ADEL | 207 | 0.2 | 0.5 |
| To be determined | 8 | <0.1 | <0.1 |
| English proficiency unknown | 32 | <0.1 | 2.9 |
| Economically disadvantaged | 61,715 | 62.0 | 60.5 |
| Not economically disadvantaged | 37,887 | 38.0 | 36.8 |
| American Indian or Alaska Native (All) | 453 | 0.5 | 0.5 |
| Asian (All) | 11,142 | 11.2 | 10.2 |
| Native Hawaiian or Other Pacific Islander (All) | 531 | 0.5 | 0.5 |
| Filipino (All) | 2,884 | 2.9 | 2.6 |
| Hispanic or Latino (All) | 56,049 | 56.3 | 53.8 |
| Black or African American (All) | 5,530 | 5.6 | 6.0 |
| White (All) | 18,516 | 18.6 | 18.9 |
| Two or more races (All) | 4,497 | 4.5 | 4.0 |
| Disability | 10,916 | 11.0 | 12.5 |
| No disability | 88,686 | 89.0 | 84.8 |
| Migrant education | 818 | 0.8 | 0.8 |
| Not migrant education | 98,784 | 99.2 | 96.5 |
| Armed forces family member | 1,406 | 1.4 | 1.2 |
| Not armed forces family member | 98,196 | 98.6 | 96.1 |
| Homeless | 3,966 | 4.0 | 4.9 |
| Not homeless | 95,636 | 96.0 | 92.3 |
| Foster youth | 506 | 0.5 | 0.6 |
| Not foster youth | 99,096 | 99.5 | 99.4 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 144 | 0.1 | 0.1 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 309 | 0.3 | 0.4 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 6,637 | 6.7 | 6.0 |
| Asian (Primary ethnicity—Economically disadvantaged) | 4,505 | 4.5 | 4.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 188 | 0.2 | 0.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 343 | 0.3 | 0.4 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 1,828 | 1.8 | 1.6 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 1,056 | 1.1 | 1.0 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 12,432 | 12.5 | 12.0 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 43,617 | 43.8 | 41.7 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 1,630 | 1.6 | 1.7 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 3,900 | 3.9 | 4.3 |
| White (Primary ethnicity—Not economically disadvantaged) | 12,303 | 12.4 | 12.4 |
| White (Primary ethnicity—Economically disadvantaged) | 6,213 | 6.2 | 6.6 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 2,725 | 2.7 | 2.4 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 1,772 | 1.8 | 1.6 |

Table 7.C.6 Demographic Student Group Summaries for High School (All Grades Tested)

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** | **Population Percent** |
| All students | 456,486 | 100.0 | 100.0 |
| Male | 233,660 | 51.2 | 50.7 |
| Female | 221,877 | 48.6 | 48.2 |
| Nonbinary | 949 | 0.2 | 0.2 |
| EL | 50,194 | 11.0 | 11.4 |
| English only | 243,514 | 53.3 | 53.5 |
| RFEP | 140,381 | 30.8 | 29.3 |
| IFEP | 22,053 | 4.8 | 4.7 |
| ADEL | 248 | <0.1 | 0.1 |
| To be determined | 34 | <0.1 | <0.1 |
| English proficiency unknown | 62 | <0.1 | 0.9 |
| Economically disadvantaged | 285,064 | 62.4 | 62.1 |
| Not economically disadvantaged | 171,422 | 37.6 | 37.0 |
| American Indian or Alaska Native (All) | 2,009 | 0.4 | 0.4 |
| Asian (All) | 46,083 | 10.1 | 9.7 |
| Native Hawaiian or Other Pacific Islander (All) | 2,026 | 0.4 | 0.5 |
| Filipino (All) | 12,069 | 2.6 | 2.5 |
| Hispanic or Latino (All) | 259,123 | 56.8 | 55.9 |
| Black or African American (All) | 21,683 | 4.7 | 5.0 |
| White (All) | 92,799 | 20.3 | 20.5 |
| Two or more races (All) | 20,694 | 4.5 | 3.9 |
| Disability | 49,603 | 10.9 | 11.8 |
| No disability | 406,883 | 89.1 | 87.3 |
| Migrant education | 3,303 | 0.7 | 0.7 |
| Not migrant education | 453,183 | 99.3 | 98.4 |
| Armed forces family member | 6,691 | 1.5 | 1.4 |
| Not armed forces family member | 449,795 | 98.5 | 97.8 |
| Homeless | 14,833 | 3.2 | 3.8 |
| Not homeless | 441,653 | 96.8 | 95.3 |
| Foster youth | 1,789 | 0.4 | 0.4 |
| Not foster youth | 454,697 | 99.6 | 99.6 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 686 | 0.2 | 0.1 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 1,323 | 0.3 | 0.3 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 26,280 | 5.8 | 5.5 |
| Asian (Primary ethnicity—Economically disadvantaged) | 19,803 | 4.3 | 4.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 731 | 0.2 | 0.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 1,295 | 0.3 | 0.3 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 7,362 | 1.6 | 1.5 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 4,707 | 1.0 | 1.0 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 57,008 | 12.5 | 12.3 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 202,115 | 44.3 | 43.6 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 6,188 | 1.4 | 1.4 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 15,495 | 3.4 | 3.6 |
| White (Primary ethnicity—Not economically disadvantaged) | 60,690 | 13.3 | 13.3 |
| White (Primary ethnicity—Economically disadvantaged) | 32,109 | 7.0 | 7.2 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 12,477 | 2.7 | 2.4 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 8,217 | 1.8 | 1.5 |

### Appendix 7.D: Demographic Student Group Summaries of Overall Achievement Levels

**Note:** Student results are suppressed and indicated as “N/A” in table 7.D.1 through table 7.D.6 where fewer than 11 students are reported in a category.

Table 7.D.1 Percent of Students in Each Achievement Level for Total Scores by Demographic Student Group for Grade Five

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Scale Score Mean** | **Scale Score SD** | **Percent in Achievement Level Standard Not Met** | **Percent in Achievement Level Standard Nearly Met** | **Percent in Achievement Level Standard Met** | **Percent in Achievement Level Standard Exceeded** | **Percent in Achievement Level Standard Met/Exceeded** |
| All students | 422,398 | 201 | 22.2 | 18.5 | 48.8 | 20.5 | 12.2 | 32.7 |
| Male | 215,552 | 201 | 22.7 | 19.9 | 46.8 | 20.5 | 12.7 | 33.2 |
| Female | 206,730 | 202 | 21.6 | 17.0 | 50.9 | 20.5 | 11.6 | 32.1 |
| Nonbinary | 116 | 205 | 23.9 | 17.2 | 44.8 | 18.1 | 19.8 | 37.9 |
| EL | 80,657 | 184 | 14.2 | 41.5 | 54.8 | 3.4 | 0.4 | 3.7 |
| English only | 259,625 | 204 | 22.1 | 15.1 | 47.5 | 23.2 | 14.2 | 37.4 |
| RFEP | 55,501 | 209 | 19.0 | 6.6 | 50.4 | 28.9 | 14.1 | 43.0 |
| IFEP | 26,483 | 213 | 20.8 | 6.0 | 41.1 | 28.5 | 24.4 | 52.9 |
| ADEL | 0 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| To be determined | 69 | 176 | 15.9 | 76.8 | 18.8 | 1.4 | 2.9 | 4.3 |
| English proficiency unknown | 63 | 201 | 22.4 | 20.6 | 49.2 | 15.9 | 14.3 | 30.2 |
| Economically disadvantaged | 277,306 | 196 | 20.3 | 23.8 | 54.2 | 15.9 | 6.1 | 22.0 |
| Not economically disadvantaged | 145,092 | 212 | 21.5 | 8.3 | 38.6 | 29.3 | 23.8 | 53.1 |
| American Indian or Alaska Native (All) | 1,807 | 195 | 20.6 | 26.1 | 52.1 | 16.4 | 5.4 | 21.7 |
| Asian (All) | 44,954 | 216 | 22.1 | 7.9 | 31.8 | 28.9 | 31.4 | 60.3 |
| Native Hawaiian or Other Pacific Islander (All) | 1,708 | 196 | 20.0 | 21.0 | 56.9 | 16.2 | 5.9 | 22.1 |
| Filipino (All) | 9,411 | 212 | 19.8 | 6.1 | 42.3 | 31.3 | 20.3 | 51.6 |
| Hispanic or Latino (All) | 234,953 | 195 | 19.9 | 23.3 | 55.6 | 15.8 | 5.3 | 21.1 |
| Black or African American (All) | 20,881 | 191 | 19.8 | 32.3 | 51.4 | 12.5 | 3.9 | 16.4 |
| White (All) | 84,422 | 210 | 21.7 | 10.3 | 41.5 | 28.3 | 19.9 | 48.3 |
| Two or more races (All) | 24,262 | 209 | 22.7 | 11.9 | 40.1 | 26.8 | 21.2 | 48.0 |
| Disability | 56,571 | 186 | 19.5 | 45.3 | 42.9 | 8.1 | 3.7 | 11.8 |
| No disability | 365,827 | 204 | 21.6 | 14.3 | 49.8 | 22.4 | 13.5 | 35.9 |
| Migrant education | 3,438 | 188 | 17.8 | 36.5 | 52.8 | 8.9 | 1.8 | 10.7 |
| Not migrant education | 418,960 | 202 | 22.2 | 18.3 | 48.8 | 20.6 | 12.3 | 32.9 |
| Armed forces family member | 6,493 | 203 | 21.2 | 15.2 | 50.0 | 22.9 | 11.8 | 34.8 |
| Not armed forces family member | 415,905 | 201 | 22.2 | 18.5 | 48.8 | 20.5 | 12.2 | 32.6 |
| Homeless | 15,571 | 190 | 19.3 | 34.3 | 51.2 | 11.1 | 3.3 | 14.4 |
| Not homeless | 406,827 | 202 | 22.2 | 17.9 | 48.7 | 20.9 | 12.5 | 33.4 |
| Foster youth | 1,355 | 187 | 18.1 | 39.6 | 49.5 | 8.9 | 2.0 | 10.8 |
| Not foster youth | 421,043 | 202 | 22.2 | 18.4 | 48.8 | 20.5 | 12.2 | 32.8 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 466 | 205 | 22.0 | 14.2 | 47.4 | 25.8 | 12.7 | 38.4 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 1,341 | 191 | 19.0 | 30.3 | 53.8 | 13.1 | 2.8 | 16.0 |
| Asian (Primary ethnicity—not economically disadvantaged) | 25,678 | 222 | 19.8 | 4.2 | 24.1 | 31.2 | 40.4 | 71.7 |
| Asian (Primary ethnicity—economically disadvantaged) | 19,276 | 208 | 22.6 | 12.8 | 42.1 | 25.8 | 19.3 | 45.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 425 | 203 | 20.6 | 12.0 | 55.1 | 20.2 | 12.7 | 32.9 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1,283 | 194 | 19.3 | 24.0 | 57.4 | 14.9 | 3.7 | 18.6 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 5,313 | 215 | 19.1 | 4.5 | 37.1 | 33.9 | 24.6 | 58.5 |
| Filipino (Primary ethnicity—economically disadvantaged) | 4,098 | 208 | 19.8 | 8.3 | 49.1 | 27.9 | 14.6 | 42.6 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 45,455 | 204 | 21.1 | 14.2 | 49.8 | 24.2 | 11.8 | 36.0 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 189,498 | 194 | 19.0 | 25.5 | 57.0 | 13.7 | 3.8 | 17.5 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 4,138 | 201 | 21.1 | 17.0 | 51.6 | 22.1 | 9.3 | 31.4 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 16,743 | 189 | 18.6 | 36.0 | 51.3 | 10.2 | 2.5 | 12.7 |
| White (Primary ethnicity—not economically disadvantaged) | 50,304 | 215 | 20.0 | 5.3 | 36.4 | 32.4 | 25.9 | 58.3 |
| White (Primary ethnicity—economically disadvantaged) | 34,118 | 202 | 21.6 | 17.5 | 49.0 | 22.3 | 11.2 | 33.5 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 13,313 | 217 | 20.6 | 5.7 | 32.6 | 31.6 | 30.1 | 61.6 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 10,949 | 201 | 21.9 | 19.5 | 49.1 | 21.0 | 10.4 | 31.4 |

Table 7.D.2 Percent of Students in Each Achievement Level for Total Scores by Demographic Student Group for Grade Eight

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Scale Score Mean** | **Scale Score SD** | **Percent in Achievement Level Standard Not Met** | **Percent in Achievement Level Standard Nearly Met** | **Percent in Achievement Level Standard Met** | **Percent in Achievement Level Standard Exceeded** | **Percent in Achievement Level Standard Met/Exceeded** |
| All students | 427,448 | 401 | 22.2 | 16.5 | 54.6 | 18.9 | 10.0 | 28.9 |
| Male | 219,022 | 401 | 22.8 | 17.7 | 52.4 | 18.8 | 11.1 | 29.9 |
| Female | 208,044 | 401 | 21.4 | 15.2 | 56.9 | 19.1 | 8.9 | 28.0 |
| Nonbinary | 382 | 410 | 21.2 | 6.5 | 49.0 | 28.0 | 16.5 | 44.5 |
| EL | 54,363 | 380 | 12.0 | 43.9 | 54.9 | 1.1 | <0.1 | 1.2 |
| English only | 243,184 | 404 | 22.2 | 13.7 | 52.7 | 21.7 | 11.9 | 33.6 |
| RFEP | 111,588 | 402 | 20.0 | 10.9 | 61.2 | 19.9 | 8.0 | 28.0 |
| IFEP | 18,220 | 416 | 21.6 | 5.2 | 38.6 | 29.1 | 27.0 | 56.1 |
| ADEL | 0 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| To be determined | 57 | 378 | 16.1 | 68.4 | 24.6 | 5.3 | 1.8 | 7.0 |
| English proficiency unknown | 36 | 394 | 20.0 | 25.0 | 61.1 | 11.1 | 2.8 | 13.9 |
| Economically disadvantaged | 278,202 | 395 | 20.0 | 21.0 | 60.1 | 14.3 | 4.6 | 18.9 |
| Not economically disadvantaged | 149,246 | 411 | 22.2 | 8.1 | 44.3 | 27.6 | 20.0 | 47.6 |
| American Indian or Alaska Native (All) | 1,806 | 394 | 20.1 | 23.3 | 58.3 | 14.3 | 4.1 | 18.4 |
| Asian (All) | 42,481 | 418 | 21.9 | 5.6 | 33.6 | 30.1 | 30.6 | 60.8 |
| Native Hawaiian or Other Pacific Islander (All) | 1,852 | 396 | 20.2 | 19.8 | 61.1 | 13.3 | 5.8 | 19.2 |
| Filipino (All) | 10,074 | 412 | 19.7 | 4.9 | 46.6 | 31.9 | 16.6 | 48.5 |
| Hispanic or Latino (All) | 242,458 | 395 | 19.5 | 20.8 | 61.4 | 13.9 | 3.8 | 17.8 |
| Black or African American (All) | 21,441 | 391 | 19.0 | 27.4 | 58.9 | 11.1 | 2.6 | 13.6 |
| White (All) | 85,485 | 409 | 22.0 | 9.4 | 47.5 | 26.6 | 16.5 | 43.1 |
| Two or more races (All) | 21,851 | 409 | 22.7 | 10.3 | 46.2 | 25.3 | 18.2 | 43.5 |
| Disability | 50,652 | 385 | 17.3 | 39.5 | 52.8 | 5.7 | 2.0 | 7.7 |
| No disability | 376,796 | 403 | 21.9 | 13.4 | 54.8 | 20.7 | 11.1 | 31.8 |
| Migrant education | 3,341 | 389 | 18.1 | 29.5 | 59.7 | 9.2 | 1.6 | 10.9 |
| Not migrant education | 424,107 | 401 | 22.2 | 16.4 | 54.6 | 19.0 | 10.1 | 29.1 |
| Armed forces family member | 5,989 | 402 | 21.7 | 14.4 | 55.0 | 20.7 | 9.9 | 30.6 |
| Not armed forces family member | 421,459 | 401 | 22.2 | 16.5 | 54.6 | 18.9 | 10.0 | 28.9 |
| Homeless | 13,775 | 390 | 18.8 | 30.1 | 57.8 | 9.9 | 2.2 | 12.1 |
| Not homeless | 413,673 | 401 | 22.2 | 16.0 | 54.5 | 19.2 | 10.3 | 29.5 |
| Foster youth | 1,431 | 386 | 17.4 | 37.5 | 54.3 | 7.2 | 1.0 | 8.2 |
| Not foster youth | 426,017 | 401 | 22.1 | 16.4 | 54.6 | 19.0 | 10.0 | 29.0 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 527 | 402 | 20.6 | 13.5 | 55.2 | 23.9 | 7.4 | 31.3 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 1,279 | 391 | 19.0 | 27.4 | 59.5 | 10.4 | 2.7 | 13.1 |
| Asian (Primary ethnicity—not economically disadvantaged) | 24,606 | 423 | 19.5 | 2.9 | 25.1 | 32.4 | 39.6 | 71.9 |
| Asian (Primary ethnicity—economically disadvantaged) | 17,875 | 410 | 22.4 | 9.3 | 45.3 | 27.0 | 18.4 | 45.4 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 518 | 405 | 21.7 | 11.8 | 53.1 | 22.4 | 12.7 | 35.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1,334 | 392 | 18.4 | 22.9 | 64.2 | 9.8 | 3.1 | 13.0 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 5,934 | 415 | 19.3 | 3.7 | 41.7 | 34.5 | 20.1 | 54.7 |
| Filipino (Primary ethnicity—economically disadvantaged) | 4,140 | 408 | 19.6 | 6.8 | 53.7 | 28.1 | 11.4 | 39.6 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 48,524 | 402 | 21.0 | 13.6 | 56.7 | 21.2 | 8.6 | 29.7 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 193,934 | 393 | 18.7 | 22.6 | 62.6 | 12.1 | 2.6 | 14.8 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 4,894 | 399 | 20.3 | 17.3 | 58.1 | 19.1 | 5.5 | 24.6 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 16,547 | 389 | 17.9 | 30.4 | 59.2 | 8.7 | 1.7 | 10.4 |
| White (Primary ethnicity—not economically disadvantaged) | 51,833 | 414 | 20.9 | 5.6 | 42.1 | 30.6 | 21.7 | 52.3 |
| White (Primary ethnicity—economically disadvantaged) | 33,652 | 401 | 21.3 | 15.3 | 55.8 | 20.3 | 8.6 | 28.9 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 12,410 | 416 | 21.3 | 5.4 | 38.4 | 30.5 | 25.8 | 56.3 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 9,441 | 400 | 21.3 | 16.8 | 56.5 | 18.5 | 8.2 | 26.7 |

Table 7.D.3 Percent of Students in Each Achievement Level for Total Scores by Demographic Student Group for Grade Ten

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Scale Score Mean** | **Scale Score SD** | **Percent in Achievement Level Standard Not Met** | **Percent in Achievement Level Standard Nearly Met** | **Percent in Achievement Level Standard Met** | **Percent in Achievement Level Standard Exceeded** | **Percent in Achievement Level Standard Met/Exceeded** |
| All students | 24,935 | 599 | 20.3 | 11.7 | 63.6 | 19.7 | 4.9 | 24.6 |
| Male | 12,811 | 598 | 21.0 | 13.5 | 61.5 | 19.6 | 5.4 | 25.0 |
| Female | 12,060 | 599 | 19.6 | 9.9 | 66.0 | 19.8 | 4.4 | 24.2 |
| Nonbinary | 64 | 608 | 21.0 | 7.8 | 42.2 | 37.5 | 12.5 | 50.0 |
| EL | 3,070 | 582 | 11.1 | 30.9 | 68.0 | 1.1 | 0.0 | 1.1 |
| English only | 13,756 | 600 | 20.5 | 10.4 | 61.6 | 22.3 | 5.6 | 27.9 |
| RFEP | 7,077 | 601 | 18.8 | 6.8 | 68.3 | 20.6 | 4.2 | 24.8 |
| IFEP | 986 | 613 | 21.5 | 4.3 | 44.3 | 34.7 | 16.7 | 51.4 |
| ADEL | 41 | 583 | 14.4 | 26.8 | 68.3 | 4.9 | 0.0 | 4.9 |
| To be determined | 0 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 5 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Economically disadvantaged | 16,713 | 595 | 18.6 | 14.0 | 68.6 | 15.1 | 2.3 | 17.4 |
| Not economically disadvantaged | 8,222 | 607 | 21.5 | 7.2 | 53.6 | 29.0 | 10.3 | 39.3 |
| American Indian or Alaska Native (All) | 129 | 598 | 21.5 | 15.5 | 59.7 | 16.3 | 8.5 | 24.8 |
| Asian (All) | 1,638 | 619 | 20.7 | 3.2 | 32.7 | 38.7 | 25.5 | 64.2 |
| Native Hawaiian or Other Pacific Islander (All) | 84 | 596 | 17.5 | 9.5 | 75.0 | 13.1 | 2.4 | 15.5 |
| Filipino (All) | 560 | 613 | 18.8 | 2.3 | 45.9 | 40.5 | 11.3 | 51.8 |
| Hispanic or Latino (All) | 15,379 | 594 | 18.0 | 13.8 | 70.1 | 14.5 | 1.7 | 16.1 |
| Black or African American (All) | 1,463 | 590 | 17.0 | 20.9 | 68.3 | 9.8 | 1.0 | 10.8 |
| White (All) | 4,674 | 605 | 20.5 | 6.9 | 56.4 | 29.2 | 7.6 | 36.8 |
| Two or more races (All) | 1,008 | 606 | 22.3 | 8.2 | 52.3 | 28.8 | 10.7 | 39.5 |
| Disability | 3,090 | 585 | 15.0 | 28.9 | 65.4 | 5.0 | 0.7 | 5.7 |
| No disability | 21,845 | 601 | 20.2 | 9.3 | 63.4 | 21.8 | 5.5 | 27.3 |
| Migrant education | 233 | 591 | 17.6 | 18.5 | 67.8 | 13.7 | 0.0 | 13.7 |
| Not migrant education | 24,702 | 599 | 20.4 | 11.7 | 63.6 | 19.8 | 5.0 | 24.7 |
| Armed forces family member | 314 | 601 | 19.9 | 8.9 | 63.1 | 23.2 | 4.8 | 28.0 |
| Not armed forces family member | 24,621 | 599 | 20.4 | 11.8 | 63.6 | 19.7 | 4.9 | 24.6 |
| Homeless | 758 | 593 | 18.0 | 16.5 | 67.9 | 14.2 | 1.3 | 15.6 |
| Not homeless | 24,177 | 599 | 20.4 | 11.6 | 63.5 | 19.9 | 5.0 | 24.9 |
| Foster youth | 150 | 588 | 18.9 | 28.7 | 57.3 | 12.7 | 1.3 | 14.0 |
| Not foster youth | 24,785 | 599 | 20.3 | 11.6 | 63.7 | 19.7 | 4.9 | 24.7 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 39 | 608 | 20.2 | 5.1 | 56.4 | 23.1 | 15.4 | 38.5 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 90 | 593 | 20.4 | 20.0 | 61.1 | 13.3 | 5.6 | 18.9 |
| Asian (Primary ethnicity—not economically disadvantaged) | 889 | 623 | 19.2 | 1.6 | 26.9 | 39.9 | 31.6 | 71.5 |
| Asian (Primary ethnicity—economically disadvantaged) | 749 | 614 | 21.3 | 5.1 | 39.5 | 37.2 | 18.2 | 55.4 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 35 | 604 | 18.8 | 5.7 | 65.7 | 22.9 | 5.7 | 28.6 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 49 | 590 | 13.8 | 12.2 | 81.6 | 6.1 | 0.0 | 6.1 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 352 | 615 | 18.7 | 2.6 | 39.8 | 45.5 | 12.2 | 57.7 |
| Filipino (Primary ethnicity—economically disadvantaged) | 208 | 610 | 18.6 | 1.9 | 56.3 | 32.2 | 9.6 | 41.8 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 3,221 | 599 | 19.4 | 9.8 | 65.4 | 21.2 | 3.5 | 24.7 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 12,158 | 593 | 17.4 | 14.8 | 71.3 | 12.7 | 1.2 | 13.9 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 364 | 595 | 19.1 | 15.7 | 65.9 | 15.4 | 3.0 | 18.4 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,099 | 588 | 15.8 | 22.7 | 69.1 | 8.0 | 0.3 | 8.3 |
| White (Primary ethnicity—not economically disadvantaged) | 2,761 | 609 | 20.6 | 5.8 | 50.2 | 33.4 | 10.7 | 44.0 |
| White (Primary ethnicity—economically disadvantaged) | 1,913 | 600 | 19.1 | 8.5 | 65.2 | 23.1 | 3.2 | 26.3 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 561 | 612 | 22.0 | 5.3 | 43.7 | 34.8 | 16.2 | 51.0 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 447 | 598 | 20.4 | 11.9 | 63.1 | 21.3 | 3.8 | 25.1 |

Table 7.D.4 Percent of Students in Each Achievement Level for Total Scores by Demographic Student Group for Grade Eleven

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Scale Score Mean** | **Scale Score SD** | **Percent in Achievement Level Standard Not Met** | **Percent in Achievement Level Standard Nearly Met** | **Percent in Achievement Level Standard Met** | **Percent in Achievement Level Standard Exceeded** | **Percent in Achievement Level Standard Met/Exceeded** |
| All students | 331,949 | 602 | 21.8 | 10.9 | 57.0 | 24.0 | 8.0 | 32.1 |
| Male | 169,911 | 602 | 22.7 | 12.8 | 54.8 | 23.4 | 9.0 | 32.4 |
| Female | 161,365 | 603 | 20.9 | 9.0 | 59.3 | 24.7 | 7.0 | 31.7 |
| Nonbinary | 673 | 610 | 21.2 | 5.8 | 48.6 | 32.8 | 12.8 | 45.6 |
| EL | 35,104 | 582 | 11.7 | 32.2 | 66.3 | 1.5 | <0.1 | 1.6 |
| English only | 178,868 | 605 | 22.0 | 9.0 | 53.8 | 27.4 | 9.9 | 37.2 |
| RFEP | 102,236 | 603 | 19.9 | 8.0 | 61.9 | 24.4 | 5.6 | 30.0 |
| IFEP | 15,690 | 615 | 21.6 | 4.1 | 41.0 | 34.4 | 20.5 | 54.9 |
| ADEL | 0 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| To be determined | 26 | 580 | 16.5 | 50.0 | 42.3 | 7.7 | 0.0 | 7.7 |
| English proficiency unknown | 25 | 604 | 20.2 | 4.0 | 68.0 | 20.0 | 8.0 | 28.0 |
| Economically disadvantaged | 206,636 | 597 | 20.0 | 13.5 | 63.8 | 18.9 | 3.7 | 22.6 |
| Not economically disadvantaged | 125,313 | 610 | 22.3 | 6.6 | 45.7 | 32.5 | 15.2 | 47.7 |
| American Indian or Alaska Native (All) | 1,427 | 598 | 20.5 | 13.5 | 62.5 | 19.3 | 4.7 | 24.0 |
| Asian (All) | 33,303 | 618 | 21.5 | 4.0 | 34.1 | 37.6 | 24.3 | 61.9 |
| Native Hawaiian or Other Pacific Islander (All) | 1,411 | 598 | 20.4 | 12.8 | 62.9 | 19.5 | 4.8 | 24.3 |
| Filipino (All) | 8,625 | 613 | 19.6 | 3.3 | 43.9 | 39.5 | 13.3 | 52.8 |
| Hispanic or Latino (All) | 187,695 | 597 | 19.5 | 13.5 | 65.1 | 18.3 | 3.0 | 21.3 |
| Black or African American (All) | 14,690 | 593 | 19.0 | 18.1 | 65.4 | 14.4 | 2.1 | 16.5 |
| White (All) | 69,609 | 609 | 22.1 | 7.2 | 47.8 | 31.9 | 13.1 | 45.0 |
| Two or more races (All) | 15,189 | 609 | 22.7 | 7.7 | 47.3 | 30.1 | 14.9 | 45.0 |
| Disability | 35,597 | 586 | 16.1 | 28.0 | 64.8 | 6.0 | 1.2 | 7.2 |
| No disability | 296,352 | 604 | 21.6 | 8.9 | 56.1 | 26.2 | 8.9 | 35.1 |
| Migrant education | 2,252 | 591 | 18.6 | 20.7 | 64.5 | 13.4 | 1.4 | 14.7 |
| Not migrant education | 329,697 | 602 | 21.8 | 10.8 | 57.0 | 24.1 | 8.1 | 32.2 |
| Armed forces family member | 4,971 | 602 | 21.1 | 9.9 | 58.7 | 24.1 | 7.2 | 31.4 |
| Not armed forces family member | 326,978 | 602 | 21.8 | 10.9 | 57.0 | 24.0 | 8.1 | 32.1 |
| Homeless | 10,109 | 591 | 18.6 | 21.0 | 64.5 | 12.9 | 1.6 | 14.5 |
| Not homeless | 321,840 | 603 | 21.8 | 10.6 | 56.8 | 24.4 | 8.2 | 32.6 |
| Foster youth | 1,133 | 588 | 17.9 | 27.0 | 62.8 | 9.4 | 0.9 | 10.2 |
| Not foster youth | 330,816 | 602 | 21.8 | 10.9 | 57.0 | 24.1 | 8.1 | 32.2 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 503 | 604 | 21.4 | 9.3 | 56.5 | 25.2 | 8.9 | 34.2 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 924 | 594 | 19.2 | 15.7 | 65.8 | 16.1 | 2.4 | 18.5 |
| Asian (Primary ethnicity—not economically disadvantaged) | 18,754 | 623 | 19.6 | 2.1 | 26.0 | 39.7 | 32.2 | 71.9 |
| Asian (Primary ethnicity—economically disadvantaged) | 14,549 | 611 | 21.8 | 6.5 | 44.6 | 34.8 | 14.1 | 49.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 508 | 602 | 21.4 | 9.3 | 59.3 | 23.4 | 8.1 | 31.5 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 903 | 595 | 19.5 | 14.8 | 64.9 | 17.3 | 3.0 | 20.3 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 5,182 | 616 | 19.2 | 2.7 | 39.0 | 42.0 | 16.3 | 58.3 |
| Filipino (Primary ethnicity—economically disadvantaged) | 3,443 | 610 | 19.6 | 4.1 | 51.4 | 35.7 | 8.8 | 44.5 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 41,355 | 602 | 20.9 | 10.5 | 58.3 | 25.2 | 6.1 | 31.2 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 146,340 | 595 | 18.8 | 14.4 | 67.1 | 16.4 | 2.1 | 18.5 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 4,194 | 599 | 20.5 | 12.4 | 60.7 | 22.4 | 4.4 | 26.8 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 10,496 | 590 | 17.7 | 20.4 | 67.2 | 11.3 | 1.2 | 12.4 |
| White (Primary ethnicity—not economically disadvantaged) | 45,626 | 613 | 21.6 | 5.1 | 42.6 | 35.7 | 16.6 | 52.3 |
| White (Primary ethnicity—economically disadvantaged) | 23,983 | 602 | 21.2 | 11.1 | 57.7 | 24.9 | 6.4 | 31.3 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 9,191 | 614 | 21.7 | 4.6 | 40.4 | 34.9 | 20.1 | 55.0 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 5,998 | 601 | 21.6 | 12.4 | 57.9 | 22.7 | 6.9 | 29.6 |

Table 7.D.5 Percent of Students in Each Achievement Level for Total Scores by Demographic Student Group for Grade Twelve

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Scale Score Mean** | **Scale Score SD** | **Percent in Achievement Level Standard Not Met** | **Percent in Achievement Level Standard Nearly Met** | **Percent in Achievement Level Standard Met** | **Percent in Achievement Level Standard Exceeded** | **Percent in Achievement Level Standard Met/Exceeded** |
| All students | 99,602 | 599 | 21.9 | 14.3 | 58.9 | 19.7 | 7.0 | 26.7 |
| Male | 50,938 | 598 | 22.7 | 16.5 | 56.3 | 19.2 | 8.0 | 27.2 |
| Female | 48,452 | 599 | 21.0 | 12.1 | 61.7 | 20.2 | 6.0 | 26.2 |
| Nonbinary | 212 | 609 | 20.0 | 6.1 | 52.8 | 29.2 | 11.8 | 41.0 |
| EL | 12,020 | 580 | 11.2 | 36.0 | 62.8 | 1.1 | <0.1 | 1.1 |
| English only | 50,890 | 602 | 22.4 | 12.4 | 56.1 | 22.9 | 8.7 | 31.6 |
| RFEP | 31,068 | 599 | 20.1 | 10.6 | 64.4 | 19.8 | 5.2 | 24.9 |
| IFEP | 5,377 | 612 | 22.4 | 5.7 | 45.1 | 31.5 | 17.7 | 49.2 |
| ADEL | 207 | 584 | 12.5 | 25.6 | 71.0 | 3.4 | 0.0 | 3.4 |
| To be determined | 8 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 32 | 591 | 18.6 | 21.9 | 65.6 | 9.4 | 3.1 | 12.5 |
| Economically disadvantaged | 61,715 | 594 | 19.5 | 17.4 | 65.0 | 14.6 | 3.0 | 17.6 |
| Not economically disadvantaged | 37,887 | 607 | 23.1 | 9.3 | 49.1 | 28.0 | 13.6 | 41.6 |
| American Indian or Alaska Native (All) | 453 | 595 | 19.9 | 16.8 | 64.2 | 15.7 | 3.3 | 19.0 |
| Asian (All) | 11,142 | 615 | 22.8 | 5.4 | 37.4 | 34.3 | 22.9 | 57.2 |
| Native Hawaiian or Other Pacific Islander (All) | 531 | 594 | 18.6 | 14.1 | 69.7 | 13.2 | 3.0 | 16.2 |
| Filipino (All) | 2,884 | 610 | 20.5 | 4.9 | 48.5 | 36.1 | 10.6 | 46.7 |
| Hispanic or Latino (All) | 56,049 | 593 | 18.9 | 17.3 | 66.5 | 13.9 | 2.3 | 16.2 |
| Black or African American (All) | 5,530 | 589 | 18.0 | 23.2 | 65.0 | 10.2 | 1.6 | 11.8 |
| White (All) | 18,516 | 606 | 22.9 | 10.4 | 50.4 | 27.4 | 11.8 | 39.2 |
| Two or more races (All) | 4,497 | 605 | 23.1 | 10.3 | 50.9 | 26.9 | 12.0 | 38.8 |
| Disability | 10,916 | 585 | 16.3 | 30.3 | 62.4 | 6.0 | 1.3 | 7.3 |
| No disability | 88,686 | 601 | 21.9 | 12.4 | 58.5 | 21.4 | 7.7 | 29.1 |
| Migrant education | 818 | 588 | 16.5 | 24.2 | 66.3 | 9.0 | 0.5 | 9.5 |
| Not migrant education | 98,784 | 599 | 21.9 | 14.3 | 58.9 | 19.8 | 7.1 | 26.9 |
| Armed forces family member | 1,406 | 599 | 21.8 | 13.9 | 59.2 | 19.5 | 7.4 | 26.9 |
| Not armed forces family member | 98,196 | 599 | 21.9 | 14.4 | 58.9 | 19.7 | 7.0 | 26.7 |
| Homeless | 3,966 | 590 | 18.0 | 22.0 | 65.9 | 10.4 | 1.6 | 12.1 |
| Not homeless | 95,636 | 599 | 22.0 | 14.0 | 58.6 | 20.1 | 7.2 | 27.3 |
| Foster youth | 506 | 584 | 15.7 | 31.8 | 60.7 | 6.9 | 0.6 | 7.5 |
| Not foster youth | 99,096 | 599 | 21.9 | 14.3 | 58.9 | 19.8 | 7.1 | 26.8 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 144 | 600 | 21.1 | 10.4 | 63.2 | 19.4 | 6.9 | 26.4 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 309 | 593 | 18.9 | 19.7 | 64.7 | 13.9 | 1.6 | 15.5 |
| Asian (Primary ethnicity—not economically disadvantaged) | 6,637 | 620 | 21.2 | 3.0 | 30.2 | 36.9 | 29.8 | 66.7 |
| Asian (Primary ethnicity—economically disadvantaged) | 4,505 | 607 | 22.8 | 8.9 | 48.0 | 30.4 | 12.7 | 43.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 188 | 598 | 20.5 | 12.8 | 66.0 | 14.9 | 6.4 | 21.3 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 343 | 592 | 17.1 | 14.9 | 71.7 | 12.2 | 1.2 | 13.4 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 1,828 | 613 | 19.9 | 3.6 | 44.2 | 40.6 | 11.7 | 52.2 |
| Filipino (Primary ethnicity—economically disadvantaged) | 1,056 | 605 | 20.8 | 7.1 | 55.9 | 28.3 | 8.7 | 37.0 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 12,432 | 598 | 20.8 | 14.2 | 61.4 | 19.5 | 4.8 | 24.4 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 43,617 | 592 | 18.1 | 18.2 | 67.9 | 12.3 | 1.6 | 13.9 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 1,630 | 595 | 20.0 | 17.2 | 63.1 | 16.3 | 3.3 | 19.6 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 3,900 | 587 | 16.6 | 25.7 | 65.8 | 7.6 | 0.9 | 8.5 |
| White (Primary ethnicity—not economically disadvantaged) | 12,303 | 609 | 22.7 | 8.0 | 46.3 | 30.8 | 14.8 | 45.6 |
| White (Primary ethnicity—economically disadvantaged) | 6,213 | 599 | 21.6 | 15.2 | 58.3 | 20.6 | 6.0 | 26.5 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,725 | 611 | 22.9 | 7.0 | 44.3 | 32.1 | 16.7 | 48.7 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,772 | 597 | 20.9 | 15.3 | 61.1 | 18.8 | 4.8 | 23.6 |

Table 7.D.6 Percent of Students in Each Achievement Level for Total Scores by Demographic Student Group for High School

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Scale Score Mean** | **Scale Score SD** | **Percent in Achievement Level Standard Not Met** | **Percent in Achievement Level Standard Nearly Met** | **Percent in Achievement Level Standard Met** | **Percent in Achievement Level Standard Exceeded** | **Percent in Achievement Level Standard Met/Exceeded** |
| All students | 456,486 | 601 | 21.8 | 11.7 | 57.8 | 22.9 | 7.6 | 30.5 |
| Male | 233,660 | 601 | 22.6 | 13.6 | 55.5 | 22.3 | 8.6 | 30.8 |
| Female | 221,877 | 602 | 20.9 | 9.7 | 60.2 | 23.4 | 6.6 | 30.1 |
| Nonbinary | 949 | 609 | 20.9 | 6.0 | 49.1 | 32.3 | 12.5 | 44.9 |
| EL | 50,194 | 581 | 11.6 | 33.0 | 65.6 | 1.4 | <0.1 | 1.4 |
| English only | 243,514 | 604 | 22.0 | 9.8 | 54.7 | 26.1 | 9.4 | 35.5 |
| RFEP | 140,381 | 602 | 19.9 | 8.5 | 62.8 | 23.2 | 5.5 | 28.6 |
| IFEP | 22,053 | 614 | 21.8 | 4.5 | 42.1 | 33.7 | 19.7 | 53.4 |
| ADEL | 248 | 584 | 12.8 | 25.8 | 70.6 | 3.6 | 0.0 | 3.6 |
| To be determined | 34 | 582 | 18.6 | 41.2 | 47.1 | 11.8 | 0.0 | 11.8 |
| English proficiency unknown | 62 | 597 | 20.0 | 14.5 | 64.5 | 16.1 | 4.8 | 21.0 |
| Economically disadvantaged | 285,064 | 596 | 19.8 | 14.4 | 64.4 | 17.8 | 3.5 | 21.2 |
| Not economically disadvantaged | 171,422 | 610 | 22.5 | 7.2 | 46.8 | 31.3 | 14.6 | 45.9 |
| American Indian or Alaska Native (All) | 2,009 | 597 | 20.5 | 14.3 | 62.7 | 18.3 | 4.6 | 22.9 |
| Asian (All) | 46,083 | 617 | 21.8 | 4.3 | 34.9 | 36.8 | 24.0 | 60.8 |
| Native Hawaiian or Other Pacific Islander (All) | 2,026 | 597 | 19.9 | 13.0 | 65.2 | 17.6 | 4.2 | 21.8 |
| Filipino (All) | 12,069 | 612 | 19.8 | 3.6 | 45.1 | 38.7 | 12.5 | 51.3 |
| Hispanic or Latino (All) | 259,123 | 596 | 19.3 | 14.4 | 65.7 | 17.1 | 2.8 | 19.9 |
| Black or African American (All) | 21,683 | 592 | 18.7 | 19.6 | 65.5 | 13.0 | 1.9 | 14.9 |
| White (All) | 92,799 | 608 | 22.2 | 7.8 | 48.7 | 30.9 | 12.6 | 43.5 |
| Two or more races (All) | 20,694 | 608 | 22.8 | 8.3 | 48.4 | 29.3 | 14.1 | 43.4 |
| Disability | 49,603 | 585 | 16.0 | 28.6 | 64.3 | 5.9 | 1.2 | 7.1 |
| No disability | 406,883 | 603 | 21.6 | 9.6 | 57.0 | 24.9 | 8.4 | 33.4 |
| Migrant education | 3,303 | 591 | 18.0 | 21.4 | 65.2 | 12.3 | 1.1 | 13.4 |
| Not migrant education | 453,183 | 601 | 21.8 | 11.6 | 57.7 | 22.9 | 7.7 | 30.6 |
| Armed forces family member | 6,691 | 601 | 21.2 | 10.7 | 59.0 | 23.1 | 7.2 | 30.3 |
| Not armed forces family member | 449,795 | 601 | 21.8 | 11.7 | 57.8 | 22.9 | 7.7 | 30.5 |
| Homeless | 14,833 | 591 | 18.4 | 21.1 | 65.1 | 12.3 | 1.6 | 13.9 |
| Not homeless | 441,653 | 602 | 21.8 | 11.4 | 57.5 | 23.2 | 7.9 | 31.1 |
| Foster youth | 1,789 | 587 | 17.5 | 28.5 | 61.7 | 8.9 | 0.8 | 9.8 |
| Not foster youth | 454,697 | 601 | 21.8 | 11.6 | 57.8 | 22.9 | 7.7 | 30.6 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 686 | 603 | 21.4 | 9.3 | 57.9 | 23.9 | 8.9 | 32.8 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 1,323 | 594 | 19.2 | 16.9 | 65.2 | 15.4 | 2.4 | 17.8 |
| Asian (Primary ethnicity—not economically disadvantaged) | 26,280 | 622 | 20.1 | 2.3 | 27.1 | 39.0 | 31.6 | 70.6 |
| Asian (Primary ethnicity—economically disadvantaged) | 19,803 | 610 | 22.1 | 7.0 | 45.2 | 33.9 | 14.0 | 47.9 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 731 | 601 | 21.1 | 10.0 | 61.3 | 21.2 | 7.5 | 28.7 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1,295 | 594 | 18.7 | 14.7 | 67.3 | 15.5 | 2.4 | 17.9 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 7,362 | 615 | 19.4 | 2.9 | 40.3 | 41.8 | 14.9 | 56.8 |
| Filipino (Primary ethnicity—economically disadvantaged) | 4,707 | 609 | 19.9 | 4.7 | 52.6 | 33.9 | 8.8 | 42.7 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 57,008 | 601 | 20.9 | 11.2 | 59.4 | 23.7 | 5.6 | 29.4 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 202,115 | 594 | 18.6 | 15.3 | 67.5 | 15.3 | 2.0 | 17.2 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 6,188 | 598 | 20.4 | 13.9 | 61.7 | 20.4 | 4.1 | 24.4 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 15,495 | 589 | 17.4 | 21.9 | 67.0 | 10.1 | 1.0 | 11.1 |
| White (Primary ethnicity—not economically disadvantaged) | 60,690 | 612 | 21.8 | 5.7 | 43.7 | 34.6 | 16.0 | 50.6 |
| White (Primary ethnicity—economically disadvantaged) | 32,109 | 601 | 21.2 | 11.7 | 58.2 | 23.9 | 6.1 | 30.1 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 12,477 | 614 | 22.1 | 5.2 | 41.4 | 34.3 | 19.2 | 53.4 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 8,217 | 600 | 21.5 | 13.0 | 58.9 | 21.8 | 6.3 | 28.1 |

### Appendix 7.E: Demographic Student Group Summaries of Domain Performance Levels

**Note:** Student results are suppressed and indicated as “N/A” in table 7.E.1 through table 7.E.18 where fewer than 11 students are reported in a category.

Table 7.E.1 Percent of Students in Each Performance Level by Demographic Student Group—Earth and Space Sciences Domain, Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 421,639 | 45.9 | 42.9 | 11.3 |
| Male | 215,161 | 45.5 | 42.1 | 12.3 |
| Female | 206,362 | 46.3 | 43.6 | 10.1 |
| Nonbinary | 116 | 44.0 | 37.1 | 19.0 |
| EL | 80,469 | 80.7 | 18.8 | 0.5 |
| English only | 259,156 | 40.5 | 46.2 | 13.3 |
| RFEP | 55,432 | 29.7 | 57.6 | 12.6 |
| IFEP | 26,450 | 26.2 | 52.5 | 21.3 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 69 | 89.9 | 7.2 | 2.9 |
| English proficiency unknown | 63 | 41.3 | 44.4 | 14.3 |
| Economically disadvantaged | 276,703 | 56.1 | 38.2 | 5.7 |
| Not economically disadvantaged | 144,936 | 26.5 | 51.7 | 21.8 |
| American Indian or Alaska Native (All) | 1,799 | 57.8 | 36.9 | 5.3 |
| Asian (All) | 44,922 | 22.8 | 49.8 | 27.3 |
| Native Hawaiian or Other Pacific Islander (All) | 1,707 | 56.9 | 38.3 | 4.8 |
| Filipino (All) | 9,400 | 25.8 | 56.7 | 17.5 |
| Hispanic or Latino (All) | 234,481 | 56.4 | 38.5 | 5.1 |
| Black or African American (All) | 20,817 | 64.2 | 31.9 | 3.9 |
| White (All) | 84,310 | 30.2 | 51.0 | 18.8 |
| Two or more races (All) | 24,203 | 32.0 | 48.3 | 19.7 |
| Disability | 56,395 | 74.8 | 21.5 | 3.7 |
| No disability | 365,244 | 41.4 | 46.1 | 12.4 |
| Migrant education | 3,429 | 70.7 | 27.6 | 1.7 |
| Not migrant education | 418,210 | 45.7 | 43.0 | 11.3 |
| Armed forces family member | 6,485 | 42.3 | 46.8 | 10.9 |
| Not armed forces family member | 415,154 | 45.9 | 42.8 | 11.3 |
| Homeless | 15,516 | 67.3 | 29.5 | 3.2 |
| Not homeless | 406,123 | 45.1 | 43.4 | 11.6 |
| Foster youth | 1,349 | 72.1 | 26.2 | 1.6 |
| Not foster youth | 420,290 | 45.8 | 42.9 | 11.3 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 464 | 39.4 | 47.2 | 13.4 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 1,335 | 64.2 | 33.3 | 2.5 |
| Asian (Primary ethnicity—not economically disadvantaged) | 25,670 | 14.3 | 50.4 | 35.3 |
| Asian (Primary ethnicity—economically disadvantaged) | 19,252 | 34.3 | 49.1 | 16.6 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 425 | 44.5 | 45.6 | 9.9 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1,282 | 61.0 | 35.9 | 3.1 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 5,310 | 20.5 | 58.4 | 21.1 |
| Filipino (Primary ethnicity—economically disadvantaged) | 4,090 | 32.7 | 54.5 | 12.8 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 45,390 | 40.1 | 48.9 | 11.0 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 189,091 | 60.3 | 36.0 | 3.7 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 4,129 | 44.1 | 46.5 | 9.4 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 16,688 | 69.2 | 28.3 | 2.5 |
| White (Primary ethnicity—not economically disadvantaged) | 50,258 | 21.0 | 54.6 | 24.3 |
| White (Primary ethnicity—economically disadvantaged) | 34,052 | 43.8 | 45.7 | 10.5 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 13,290 | 20.0 | 52.3 | 27.7 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 10,913 | 46.6 | 43.4 | 10.0 |

Table 7.E.2 Percent of Students in Each Performance Level by Demographic Student Group—Earth and Space Sciences Domain, Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 426,195 | 39.9 | 48.4 | 11.8 |
| Male | 218,405 | 38.9 | 48.0 | 13.1 |
| Female | 207,409 | 40.9 | 48.8 | 10.4 |
| Nonbinary | 381 | 26.2 | 52.2 | 21.5 |
| EL | 54,126 | 75.0 | 24.6 | 0.4 |
| English only | 242,470 | 35.3 | 50.8 | 13.9 |
| RFEP | 111,320 | 36.1 | 53.7 | 10.2 |
| IFEP | 18,186 | 19.1 | 53.0 | 27.9 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 57 | 78.9 | 19.3 | 1.8 |
| English proficiency unknown | 36 | 44.4 | 50.0 | 5.6 |
| Economically disadvantaged | 277,210 | 48.2 | 45.5 | 6.3 |
| Not economically disadvantaged | 148,985 | 24.3 | 53.7 | 21.9 |
| American Indian or Alaska Native (All) | 1,795 | 49.0 | 45.6 | 5.5 |
| Asian (All) | 42,428 | 17.8 | 52.0 | 30.2 |
| Native Hawaiian or Other Pacific Islander (All) | 1,847 | 48.9 | 43.9 | 7.3 |
| Filipino (All) | 10,067 | 21.5 | 59.4 | 19.2 |
| Hispanic or Latino (All) | 241,629 | 48.6 | 45.7 | 5.7 |
| Black or African American (All) | 21,313 | 56.1 | 40.1 | 3.8 |
| White (All) | 85,310 | 26.8 | 54.0 | 19.2 |
| Two or more races (All) | 21,806 | 27.9 | 52.1 | 20.0 |
| Disability | 50,368 | 65.8 | 31.2 | 2.9 |
| No disability | 375,827 | 36.4 | 50.6 | 13.0 |
| Migrant education | 3,332 | 57.9 | 38.9 | 3.2 |
| Not migrant education | 422,863 | 39.7 | 48.4 | 11.9 |
| Armed forces family member | 5,980 | 38.8 | 49.2 | 12.0 |
| Not armed forces family member | 420,215 | 39.9 | 48.3 | 11.8 |
| Homeless | 13,691 | 56.9 | 39.1 | 3.9 |
| Not homeless | 412,504 | 39.3 | 48.7 | 12.1 |
| Foster youth | 1,415 | 64.7 | 33.0 | 2.3 |
| Not foster youth | 424,780 | 39.8 | 48.4 | 11.8 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 526 | 34.6 | 55.3 | 10.1 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 1,269 | 54.9 | 41.5 | 3.5 |
| Asian (Primary ethnicity—not economically disadvantaged) | 24,583 | 10.8 | 50.9 | 38.3 |
| Asian (Primary ethnicity—economically disadvantaged) | 17,845 | 27.4 | 53.4 | 19.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 517 | 32.5 | 50.9 | 16.6 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1,330 | 55.3 | 41.1 | 3.6 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 5,930 | 17.7 | 60.1 | 22.2 |
| Filipino (Primary ethnicity—economically disadvantaged) | 4,137 | 26.9 | 58.3 | 14.9 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 48,397 | 36.6 | 52.4 | 11.0 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 193,232 | 51.6 | 44.1 | 4.3 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 4,882 | 42.0 | 50.8 | 7.2 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 16,431 | 60.3 | 36.9 | 2.8 |
| White (Primary ethnicity—not economically disadvantaged) | 51,754 | 19.7 | 55.8 | 24.6 |
| White (Primary ethnicity—economically disadvantaged) | 33,556 | 37.9 | 51.2 | 11.0 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 12,396 | 18.4 | 54.0 | 27.6 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 9,410 | 40.5 | 49.5 | 10.0 |

Table 7.E.3 Percent of Students in Each Performance Level by Demographic Student Group—Earth and Space Sciences Domain, Grade Ten

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 24,835 | 54.4 | 35.4 | 10.2 |
| Male | 12,756 | 53.8 | 34.7 | 11.5 |
| Female | 12,015 | 55.2 | 36.0 | 8.8 |
| Nonbinary | 64 | 31.3 | 51.6 | 17.2 |
| EL | 3,050 | 89.4 | 10.4 | 0.2 |
| English only | 13,697 | 50.1 | 38.0 | 11.9 |
| RFEP | 7,057 | 50.9 | 40.1 | 9.0 |
| IFEP | 985 | 29.4 | 43.8 | 26.8 |
| ADEL | 41 | 87.8 | 9.8 | 2.4 |
| To be determined | 0 | N/A | N/A | N/A |
| English proficiency unknown | 5 | N/A | N/A | N/A |
| Economically disadvantaged | 16,630 | 61.9 | 31.8 | 6.2 |
| Not economically disadvantaged | 8,205 | 39.1 | 42.6 | 18.3 |
| American Indian or Alaska Native (All) | 128 | 50.8 | 35.9 | 13.3 |
| Asian (All) | 1,632 | 21.4 | 44.8 | 33.8 |
| Native Hawaiian or Other Pacific Islander (All) | 83 | 60.2 | 34.9 | 4.8 |
| Filipino (All) | 560 | 28.4 | 49.6 | 22.0 |
| Hispanic or Latino (All) | 15,318 | 62.8 | 31.9 | 5.4 |
| Black or African American (All) | 1,449 | 72.5 | 25.0 | 2.6 |
| White (All) | 4,661 | 39.0 | 44.5 | 16.6 |
| Two or more races (All) | 1,004 | 41.0 | 38.4 | 20.5 |
| Disability | 3,068 | 80.9 | 16.8 | 2.4 |
| No disability | 21,767 | 50.7 | 38.0 | 11.3 |
| Migrant education | 233 | 70.8 | 24.0 | 5.2 |
| Not migrant education | 24,602 | 54.3 | 35.5 | 10.3 |
| Armed forces family member | 314 | 49.4 | 38.9 | 11.8 |
| Not armed forces family member | 24,521 | 54.5 | 35.3 | 10.2 |
| Homeless | 757 | 65.1 | 30.4 | 4.5 |
| Not homeless | 24,078 | 54.1 | 35.5 | 10.4 |
| Foster youth | 147 | 72.8 | 22.4 | 4.8 |
| Not foster youth | 24,688 | 54.3 | 35.4 | 10.2 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 39 | 35.9 | 46.2 | 17.9 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 89 | 57.3 | 31.5 | 11.2 |
| Asian (Primary ethnicity—not economically disadvantaged) | 888 | 16.6 | 42.1 | 41.3 |
| Asian (Primary ethnicity—economically disadvantaged) | 744 | 27.2 | 48.0 | 24.9 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 35 | 37.1 | 54.3 | 8.6 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 48 | 77.1 | 20.8 | 2.1 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 352 | 24.7 | 52.6 | 22.7 |
| Filipino (Primary ethnicity—economically disadvantaged) | 208 | 34.6 | 44.7 | 20.7 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 3,217 | 52.3 | 38.4 | 9.3 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 12,101 | 65.5 | 30.1 | 4.4 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 362 | 62.2 | 32.3 | 5.5 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,087 | 75.9 | 22.5 | 1.6 |
| White (Primary ethnicity—not economically disadvantaged) | 2,753 | 31.9 | 47.4 | 20.7 |
| White (Primary ethnicity—economically disadvantaged) | 1,908 | 49.2 | 40.3 | 10.5 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 559 | 29.9 | 42.6 | 27.5 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 445 | 55.1 | 33.3 | 11.7 |

Table 7.E.4 Percent of Students in Each Performance Level by Demographic Student Group—Earth and Space Sciences Domain, Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 331,021 | 47.9 | 37.4 | 14.7 |
| Male | 169,446 | 47.8 | 35.9 | 16.3 |
| Female | 160,903 | 48.0 | 39.0 | 13.0 |
| Nonbinary | 672 | 33.0 | 39.7 | 27.2 |
| EL | 34,919 | 89.3 | 10.5 | 0.2 |
| English only | 178,433 | 42.0 | 39.8 | 18.2 |
| RFEP | 101,959 | 47.2 | 41.5 | 11.3 |
| IFEP | 15,662 | 27.4 | 43.1 | 29.5 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 23 | 87.0 | 13.0 | 0.0 |
| English proficiency unknown | 25 | 48.0 | 48.0 | 4.0 |
| Economically disadvantaged | 205,945 | 56.9 | 34.6 | 8.5 |
| Not economically disadvantaged | 125,076 | 33.0 | 42.0 | 25.1 |
| American Indian or Alaska Native (All) | 1,423 | 54.8 | 34.9 | 10.3 |
| Asian (All) | 33,247 | 23.6 | 43.1 | 33.3 |
| Native Hawaiian or Other Pacific Islander (All) | 1,404 | 56.9 | 33.5 | 9.5 |
| Filipino (All) | 8,613 | 27.2 | 49.1 | 23.7 |
| Hispanic or Latino (All) | 187,065 | 57.6 | 34.7 | 7.7 |
| Black or African American (All) | 14,610 | 65.6 | 28.7 | 5.7 |
| White (All) | 69,498 | 34.3 | 42.0 | 23.8 |
| Two or more races (All) | 15,161 | 36.1 | 39.6 | 24.3 |
| Disability | 35,433 | 79.3 | 17.6 | 3.1 |
| No disability | 295,588 | 44.1 | 39.8 | 16.1 |
| Migrant education | 2,242 | 68.5 | 26.8 | 4.7 |
| Not migrant education | 328,779 | 47.7 | 37.5 | 14.8 |
| Armed forces family member | 4,968 | 48.1 | 38.2 | 13.6 |
| Not armed forces family member | 326,053 | 47.9 | 37.4 | 14.8 |
| Homeless | 10,052 | 68.2 | 27.0 | 4.9 |
| Not homeless | 320,969 | 47.2 | 37.7 | 15.1 |
| Foster youth | 1,123 | 73.6 | 22.7 | 3.7 |
| Not foster youth | 329,898 | 47.8 | 37.4 | 14.8 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 502 | 43.4 | 39.4 | 17.1 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 921 | 61.0 | 32.4 | 6.6 |
| Asian (Primary ethnicity—not economically disadvantaged) | 18,730 | 15.9 | 42.1 | 42.0 |
| Asian (Primary ethnicity—economically disadvantaged) | 14,517 | 33.5 | 44.4 | 22.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 506 | 47.4 | 38.5 | 14.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 898 | 62.2 | 30.7 | 7.0 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 5,177 | 22.7 | 49.8 | 27.4 |
| Filipino (Primary ethnicity—economically disadvantaged) | 3,436 | 33.9 | 48.0 | 18.1 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 41,241 | 46.8 | 39.7 | 13.4 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 145,824 | 60.7 | 33.2 | 6.1 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 4,181 | 52.9 | 37.0 | 10.1 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 10,429 | 70.7 | 25.3 | 4.0 |
| White (Primary ethnicity—not economically disadvantaged) | 45,561 | 27.7 | 43.6 | 28.7 |
| White (Primary ethnicity—economically disadvantaged) | 23,937 | 46.8 | 38.9 | 14.3 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 9,178 | 27.0 | 42.1 | 30.9 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 5,983 | 50.0 | 35.9 | 14.1 |

Table 7.E.5 Percent of Students in Each Performance Level by Demographic Student Group—Earth and Space Sciences Domain, Grade Twelve

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 99,215 | 54.2 | 32.8 | 13.0 |
| Male | 50,728 | 54.1 | 31.4 | 14.5 |
| Female | 48,276 | 54.5 | 34.1 | 11.4 |
| Nonbinary | 211 | 30.8 | 46.4 | 22.7 |
| EL | 11,949 | 91.3 | 8.6 | 0.1 |
| English only | 50,708 | 48.2 | 35.5 | 16.3 |
| RFEP | 30,943 | 53.3 | 36.5 | 10.2 |
| IFEP | 5,368 | 32.6 | 40.7 | 26.7 |
| ADEL | 207 | 89.9 | 9.7 | 0.5 |
| To be determined | 8 | N/A | N/A | N/A |
| English proficiency unknown | 32 | 56.3 | 40.6 | 3.1 |
| Economically disadvantaged | 61,438 | 63.6 | 29.4 | 7.1 |
| Not economically disadvantaged | 37,777 | 39.0 | 38.3 | 22.7 |
| American Indian or Alaska Native (All) | 449 | 59.7 | 31.2 | 9.1 |
| Asian (All) | 11,111 | 28.3 | 39.7 | 32.0 |
| Native Hawaiian or Other Pacific Islander (All) | 530 | 63.8 | 30.6 | 5.7 |
| Filipino (All) | 2,878 | 32.9 | 46.4 | 20.7 |
| Hispanic or Latino (All) | 55,801 | 64.3 | 29.5 | 6.2 |
| Black or African American (All) | 5,493 | 71.9 | 23.9 | 4.2 |
| White (All) | 18,469 | 40.1 | 38.2 | 21.7 |
| Two or more races (All) | 4,484 | 42.5 | 36.5 | 21.0 |
| Disability | 10,832 | 80.2 | 16.3 | 3.5 |
| No disability | 88,383 | 51.0 | 34.8 | 14.2 |
| Migrant education | 814 | 74.4 | 22.7 | 2.8 |
| Not migrant education | 98,401 | 54.1 | 32.9 | 13.1 |
| Armed forces family member | 1,405 | 52.8 | 35.1 | 12.1 |
| Not armed forces family member | 97,810 | 54.3 | 32.7 | 13.0 |
| Homeless | 3,943 | 72.1 | 23.3 | 4.6 |
| Not homeless | 95,272 | 53.5 | 33.2 | 13.3 |
| Foster youth | 495 | 81.4 | 17.0 | 1.6 |
| Not foster youth | 98,720 | 54.1 | 32.9 | 13.1 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 143 | 47.6 | 39.9 | 12.6 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 306 | 65.4 | 27.1 | 7.5 |
| Asian (Primary ethnicity—not economically disadvantaged) | 6,619 | 19.9 | 40.0 | 40.0 |
| Asian (Primary ethnicity—economically disadvantaged) | 4,492 | 40.6 | 39.3 | 20.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 188 | 56.4 | 35.6 | 8.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 342 | 67.8 | 27.8 | 4.4 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 1,827 | 28.2 | 49.1 | 22.7 |
| Filipino (Primary ethnicity—economically disadvantaged) | 1,051 | 41.2 | 41.6 | 17.2 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 12,378 | 54.3 | 34.6 | 11.1 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 43,423 | 67.1 | 28.0 | 4.8 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 1,627 | 61.6 | 30.4 | 8.0 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 3,866 | 76.2 | 21.2 | 2.7 |
| White (Primary ethnicity—not economically disadvantaged) | 12,275 | 33.5 | 40.3 | 26.1 |
| White (Primary ethnicity—economically disadvantaged) | 6,194 | 53.0 | 34.0 | 13.0 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,720 | 32.7 | 40.0 | 27.4 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,764 | 57.7 | 31.1 | 11.2 |

Table 7.E.6 Percent of Students in Each Performance Level by Demographic Student Group—Earth and Space Sciences Domain, High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 455,071 | 49.6 | 36.3 | 14.1 |
| Male | 232,930 | 49.5 | 34.8 | 15.7 |
| Female | 221,194 | 49.8 | 37.8 | 12.4 |
| Nonbinary | 947 | 32.4 | 42.0 | 25.6 |
| EL | 49,918 | 89.8 | 10.0 | 0.2 |
| English only | 242,838 | 43.7 | 38.8 | 17.5 |
| RFEP | 139,959 | 48.7 | 40.3 | 11.0 |
| IFEP | 22,015 | 28.8 | 42.6 | 28.7 |
| ADEL | 248 | 89.5 | 9.7 | 0.8 |
| To be determined | 31 | 83.9 | 12.9 | 3.2 |
| English proficiency unknown | 62 | 53.2 | 43.5 | 3.2 |
| Economically disadvantaged | 284,013 | 58.7 | 33.3 | 8.0 |
| Not economically disadvantaged | 171,058 | 34.6 | 41.2 | 24.2 |
| American Indian or Alaska Native (All) | 2,000 | 55.7 | 34.1 | 10.3 |
| Asian (All) | 45,990 | 24.6 | 42.4 | 33.0 |
| Native Hawaiian or Other Pacific Islander (All) | 2,017 | 58.8 | 32.8 | 8.3 |
| Filipino (All) | 12,051 | 28.6 | 48.5 | 22.9 |
| Hispanic or Latino (All) | 258,184 | 59.4 | 33.4 | 7.2 |
| Black or African American (All) | 21,552 | 67.7 | 27.2 | 5.1 |
| White (All) | 92,628 | 35.7 | 41.3 | 23.0 |
| Two or more races (All) | 20,649 | 37.7 | 38.9 | 23.4 |
| Disability | 49,333 | 79.6 | 17.2 | 3.1 |
| No disability | 405,738 | 46.0 | 38.6 | 15.5 |
| Migrant education | 3,289 | 70.1 | 25.6 | 4.3 |
| Not migrant education | 451,782 | 49.5 | 36.3 | 14.2 |
| Armed forces family member | 6,687 | 49.2 | 37.6 | 13.2 |
| Not armed forces family member | 448,384 | 49.6 | 36.2 | 14.1 |
| Homeless | 14,752 | 69.0 | 26.2 | 4.8 |
| Not homeless | 440,319 | 49.0 | 36.6 | 14.4 |
| Foster youth | 1,765 | 75.7 | 21.1 | 3.2 |
| Not foster youth | 453,306 | 49.5 | 36.3 | 14.2 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 684 | 43.9 | 39.9 | 16.2 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 1,316 | 61.8 | 31.1 | 7.1 |
| Asian (Primary ethnicity—not economically disadvantaged) | 26,237 | 16.9 | 41.6 | 41.5 |
| Asian (Primary ethnicity—economically disadvantaged) | 19,753 | 34.9 | 43.4 | 21.7 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 729 | 49.2 | 38.5 | 12.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1,288 | 64.3 | 29.6 | 6.1 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 7,356 | 24.2 | 49.8 | 26.0 |
| Filipino (Primary ethnicity—economically disadvantaged) | 4,695 | 35.6 | 46.4 | 18.0 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 56,836 | 48.8 | 38.5 | 12.7 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 201,348 | 62.4 | 31.9 | 5.7 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 6,170 | 55.8 | 35.0 | 9.3 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 15,382 | 72.5 | 24.1 | 3.5 |
| White (Primary ethnicity—not economically disadvantaged) | 60,589 | 29.1 | 43.1 | 27.8 |
| White (Primary ethnicity—economically disadvantaged) | 32,039 | 48.1 | 38.0 | 13.8 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 12,457 | 28.4 | 41.6 | 30.0 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 8,192 | 52.0 | 34.7 | 13.3 |

Table 7.E.7 Percent of Students in Each Performance Level by Demographic Student Group—Life Sciences Domain, Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 421,655 | 49.6 | 38.0 | 12.3 |
| Male | 215,176 | 50.9 | 37.6 | 11.5 |
| Female | 206,363 | 48.3 | 38.5 | 13.2 |
| Nonbinary | 116 | 48.3 | 37.1 | 14.7 |
| EL | 80,473 | 84.4 | 14.9 | 0.6 |
| English only | 259,158 | 44.5 | 41.3 | 14.2 |
| RFEP | 55,443 | 32.9 | 51.9 | 15.2 |
| IFEP | 26,449 | 28.8 | 47.7 | 23.6 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 69 | 95.7 | 1.4 | 2.9 |
| English proficiency unknown | 63 | 49.2 | 33.3 | 17.5 |
| Economically disadvantaged | 276,726 | 59.8 | 33.1 | 7.0 |
| Not economically disadvantaged | 144,929 | 30.2 | 47.4 | 22.4 |
| American Indian or Alaska Native (All) | 1,798 | 61.0 | 32.1 | 7.0 |
| Asian (All) | 44,921 | 25.8 | 45.5 | 28.7 |
| Native Hawaiian or Other Pacific Islander (All) | 1,706 | 58.0 | 34.3 | 7.7 |
| Filipino (All) | 9,395 | 28.8 | 50.9 | 20.3 |
| Hispanic or Latino (All) | 234,508 | 60.2 | 33.4 | 6.5 |
| Black or African American (All) | 20,818 | 68.3 | 26.9 | 4.9 |
| White (All) | 84,312 | 34.4 | 46.8 | 18.8 |
| Two or more races (All) | 24,197 | 35.5 | 44.5 | 20.0 |
| Disability | 56,394 | 77.9 | 18.7 | 3.3 |
| No disability | 365,261 | 45.3 | 41.0 | 13.7 |
| Migrant education | 3,432 | 73.8 | 23.6 | 2.6 |
| Not migrant education | 418,223 | 49.4 | 38.2 | 12.4 |
| Armed forces family member | 6,487 | 46.0 | 41.2 | 12.8 |
| Not armed forces family member | 415,168 | 49.7 | 38.0 | 12.3 |
| Homeless | 15,526 | 70.2 | 25.6 | 4.2 |
| Not homeless | 406,129 | 48.8 | 38.5 | 12.6 |
| Foster youth | 1,350 | 75.5 | 21.6 | 3.0 |
| Not foster youth | 420,305 | 49.5 | 38.1 | 12.4 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 464 | 42.9 | 42.9 | 14.2 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 1,334 | 67.2 | 28.3 | 4.4 |
| Asian (Primary ethnicity—not economically disadvantaged) | 25,669 | 16.5 | 47.3 | 36.2 |
| Asian (Primary ethnicity—economically disadvantaged) | 19,252 | 38.1 | 43.1 | 18.7 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 425 | 47.3 | 37.6 | 15.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1,281 | 61.5 | 33.2 | 5.3 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 5,308 | 23.2 | 52.9 | 23.9 |
| Filipino (Primary ethnicity—economically disadvantaged) | 4,087 | 36.1 | 48.4 | 15.5 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 45,389 | 44.4 | 43.0 | 12.6 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 189,119 | 64.0 | 31.0 | 5.0 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 4,128 | 49.5 | 39.9 | 10.6 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 16,690 | 72.9 | 23.6 | 3.5 |
| White (Primary ethnicity—not economically disadvantaged) | 50,261 | 25.1 | 51.0 | 23.9 |
| White (Primary ethnicity—economically disadvantaged) | 34,051 | 48.1 | 40.5 | 11.4 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 13,285 | 22.9 | 49.7 | 27.4 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 10,912 | 50.8 | 38.3 | 10.9 |

Table 7.E.8 Percent of Students in Each Performance Level by Demographic Student Group—Life Sciences Domain, Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 426,134 | 51.7 | 36.2 | 12.1 |
| Male | 218,376 | 52.8 | 34.6 | 12.6 |
| Female | 207,378 | 50.7 | 37.8 | 11.5 |
| Nonbinary | 380 | 29.7 | 51.1 | 19.2 |
| EL | 54,117 | 92.4 | 7.4 | 0.2 |
| English only | 242,438 | 46.5 | 39.3 | 14.1 |
| RFEP | 111,301 | 47.6 | 42.0 | 10.4 |
| IFEP | 18,185 | 25.4 | 44.8 | 29.7 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 57 | 89.5 | 8.8 | 1.8 |
| English proficiency unknown | 36 | 72.2 | 25.0 | 2.8 |
| Economically disadvantaged | 277,168 | 61.6 | 31.9 | 6.5 |
| Not economically disadvantaged | 148,966 | 33.3 | 44.2 | 22.5 |
| American Indian or Alaska Native (All) | 1,795 | 63.3 | 30.2 | 6.5 |
| Asian (All) | 42,425 | 23.4 | 43.6 | 33.0 |
| Native Hawaiian or Other Pacific Islander (All) | 1,846 | 61.1 | 30.9 | 8.0 |
| Filipino (All) | 10,065 | 28.7 | 50.3 | 21.0 |
| Hispanic or Latino (All) | 241,595 | 62.3 | 32.0 | 5.7 |
| Black or African American (All) | 21,309 | 69.7 | 26.3 | 4.1 |
| White (All) | 85,294 | 37.5 | 44.0 | 18.6 |
| Two or more races (All) | 21,805 | 37.2 | 42.4 | 20.4 |
| Disability | 50,348 | 82.1 | 15.3 | 2.6 |
| No disability | 375,786 | 47.7 | 39.0 | 13.3 |
| Migrant education | 3,332 | 73.4 | 23.6 | 3.1 |
| Not migrant education | 422,802 | 51.6 | 36.3 | 12.1 |
| Armed forces family member | 5,980 | 50.1 | 37.6 | 12.4 |
| Not armed forces family member | 420,154 | 51.8 | 36.2 | 12.0 |
| Homeless | 13,687 | 71.5 | 25.0 | 3.6 |
| Not homeless | 412,447 | 51.1 | 36.6 | 12.3 |
| Foster youth | 1,414 | 77.2 | 20.7 | 2.1 |
| Not foster youth | 424,720 | 51.7 | 36.3 | 12.1 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 526 | 50.2 | 37.8 | 12.0 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 1,269 | 68.7 | 27.0 | 4.3 |
| Asian (Primary ethnicity—not economically disadvantaged) | 24,583 | 14.6 | 43.7 | 41.6 |
| Asian (Primary ethnicity—economically disadvantaged) | 17,842 | 35.4 | 43.4 | 21.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 517 | 42.7 | 41.2 | 16.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1,329 | 68.2 | 26.9 | 4.9 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 5,928 | 23.8 | 51.3 | 24.8 |
| Filipino (Primary ethnicity—economically disadvantaged) | 4,137 | 35.7 | 48.9 | 15.5 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 48,391 | 48.6 | 40.2 | 11.2 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 193,204 | 65.7 | 29.9 | 4.3 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 4,881 | 54.1 | 38.0 | 7.9 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 16,428 | 74.3 | 22.8 | 2.9 |
| White (Primary ethnicity—not economically disadvantaged) | 51,744 | 28.7 | 47.5 | 23.8 |
| White (Primary ethnicity—economically disadvantaged) | 33,550 | 51.0 | 38.4 | 10.6 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 12,396 | 25.4 | 46.5 | 28.1 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 9,409 | 52.6 | 37.1 | 10.3 |

Table 7.E.9 Percent of Students in Each Performance Level by Demographic Student Group—Life Sciences Domain, Grade Ten

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 24,809 | 50.2 | 38.6 | 11.2 |
| Male | 12,741 | 52.5 | 36.4 | 11.0 |
| Female | 12,004 | 47.8 | 40.9 | 11.3 |
| Nonbinary | 64 | 34.4 | 43.8 | 21.9 |
| EL | 3,048 | 86.8 | 13.1 | 0.2 |
| English only | 13,681 | 46.7 | 40.8 | 12.5 |
| RFEP | 7,049 | 44.2 | 44.9 | 10.9 |
| IFEP | 985 | 26.6 | 42.8 | 30.6 |
| ADEL | 41 | 80.5 | 19.5 | 0.0 |
| To be determined | 0 | N/A | N/A | N/A |
| English proficiency unknown | 5 | N/A | N/A | N/A |
| Economically disadvantaged | 16,609 | 56.9 | 36.2 | 6.9 |
| Not economically disadvantaged | 8,200 | 36.5 | 43.5 | 20.0 |
| American Indian or Alaska Native (All) | 126 | 57.9 | 27.0 | 15.1 |
| Asian (All) | 1,631 | 18.7 | 41.6 | 39.7 |
| Native Hawaiian or Other Pacific Islander (All) | 83 | 51.8 | 43.4 | 4.8 |
| Filipino (All) | 560 | 22.5 | 52.0 | 25.5 |
| Hispanic or Latino (All) | 15,301 | 57.7 | 36.1 | 6.1 |
| Black or African American (All) | 1,447 | 66.6 | 29.8 | 3.7 |
| White (All) | 4,657 | 36.7 | 46.6 | 16.7 |
| Two or more races (All) | 1,004 | 38.9 | 40.7 | 20.3 |
| Disability | 3,064 | 80.3 | 17.7 | 2.0 |
| No disability | 21,745 | 45.9 | 41.5 | 12.5 |
| Migrant education | 233 | 62.2 | 33.0 | 4.7 |
| Not migrant education | 24,576 | 50.1 | 38.7 | 11.3 |
| Armed forces family member | 314 | 44.6 | 43.9 | 11.5 |
| Not armed forces family member | 24,495 | 50.2 | 38.5 | 11.2 |
| Homeless | 756 | 60.1 | 34.0 | 6.0 |
| Not homeless | 24,053 | 49.9 | 38.8 | 11.4 |
| Foster youth | 147 | 72.1 | 21.1 | 6.8 |
| Not foster youth | 24,662 | 50.0 | 38.7 | 11.2 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 39 | 38.5 | 38.5 | 23.1 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 87 | 66.7 | 21.8 | 11.5 |
| Asian (Primary ethnicity—not economically disadvantaged) | 888 | 14.0 | 40.1 | 45.9 |
| Asian (Primary ethnicity—economically disadvantaged) | 743 | 24.4 | 43.3 | 32.3 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 35 | 34.3 | 60.0 | 5.7 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 48 | 64.6 | 31.3 | 4.2 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 352 | 19.6 | 51.7 | 28.7 |
| Filipino (Primary ethnicity—economically disadvantaged) | 208 | 27.4 | 52.4 | 20.2 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 3,216 | 48.8 | 40.5 | 10.6 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 12,085 | 60.1 | 35.0 | 4.9 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 362 | 54.4 | 37.8 | 7.7 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,085 | 70.6 | 27.1 | 2.3 |
| White (Primary ethnicity—not economically disadvantaged) | 2,749 | 30.6 | 47.8 | 21.6 |
| White (Primary ethnicity—economically disadvantaged) | 1,908 | 45.6 | 44.8 | 9.5 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 559 | 30.1 | 42.9 | 27.0 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 445 | 50.1 | 38.0 | 11.9 |

Table 7.E.10 Percent of Students in Each Performance Level by Demographic Student Group—Life Sciences Domain, Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 330,785 | 44.6 | 39.6 | 15.8 |
| Male | 169,328 | 47.4 | 36.9 | 15.7 |
| Female | 160,787 | 41.7 | 42.5 | 15.8 |
| Nonbinary | 670 | 30.7 | 44.2 | 25.1 |
| EL | 34,883 | 86.5 | 13.2 | 0.3 |
| English only | 178,316 | 39.7 | 41.6 | 18.7 |
| RFEP | 101,880 | 41.9 | 44.6 | 13.5 |
| IFEP | 15,658 | 24.2 | 43.6 | 32.2 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 23 | 87.0 | 8.7 | 4.3 |
| English proficiency unknown | 25 | 44.0 | 36.0 | 20.0 |
| Economically disadvantaged | 205,777 | 52.8 | 37.5 | 9.6 |
| Not economically disadvantaged | 125,008 | 31.0 | 43.0 | 26.0 |
| American Indian or Alaska Native (All) | 1,423 | 52.7 | 36.1 | 11.2 |
| Asian (All) | 33,230 | 20.8 | 41.9 | 37.3 |
| Native Hawaiian or Other Pacific Islander (All) | 1,402 | 53.4 | 36.5 | 10.1 |
| Filipino (All) | 8,606 | 23.5 | 50.5 | 26.0 |
| Hispanic or Latino (All) | 186,928 | 53.5 | 37.8 | 8.7 |
| Black or African American (All) | 14,590 | 61.3 | 32.2 | 6.4 |
| White (All) | 69,456 | 33.1 | 43.4 | 23.5 |
| Two or more races (All) | 15,150 | 33.6 | 41.5 | 24.9 |
| Disability | 35,383 | 77.9 | 19.5 | 2.7 |
| No disability | 295,402 | 40.6 | 42.0 | 17.4 |
| Migrant education | 2,239 | 64.4 | 30.9 | 4.7 |
| Not migrant education | 328,546 | 44.5 | 39.7 | 15.9 |
| Armed forces family member | 4,963 | 44.8 | 40.9 | 14.3 |
| Not armed forces family member | 325,822 | 44.6 | 39.6 | 15.8 |
| Homeless | 10,041 | 64.2 | 30.3 | 5.5 |
| Not homeless | 320,744 | 44.0 | 39.9 | 16.1 |
| Foster youth | 1,119 | 71.0 | 25.3 | 3.8 |
| Not foster youth | 329,666 | 44.5 | 39.7 | 15.8 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 502 | 39.6 | 43.0 | 17.3 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 921 | 59.8 | 32.2 | 7.9 |
| Asian (Primary ethnicity—not economically disadvantaged) | 18,726 | 13.9 | 40.1 | 46.0 |
| Asian (Primary ethnicity—economically disadvantaged) | 14,504 | 29.7 | 44.2 | 26.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 506 | 45.3 | 40.5 | 14.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 896 | 58.0 | 34.3 | 7.7 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 5,176 | 19.8 | 50.9 | 29.4 |
| Filipino (Primary ethnicity—economically disadvantaged) | 3,430 | 29.2 | 49.9 | 20.9 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 41,214 | 43.9 | 42.1 | 14.0 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 145,714 | 56.2 | 36.6 | 7.2 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 4,176 | 48.8 | 39.7 | 11.6 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 10,414 | 66.4 | 29.2 | 4.4 |
| White (Primary ethnicity—not economically disadvantaged) | 45,535 | 27.0 | 44.5 | 28.5 |
| White (Primary ethnicity—economically disadvantaged) | 23,921 | 44.6 | 41.3 | 14.1 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 9,173 | 24.6 | 43.3 | 32.1 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 5,977 | 47.4 | 38.7 | 13.9 |

Table 7.E.11 Percent of Students in Each Performance Level by Demographic Student Group—Life Sciences Domain, Grade Twelve

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 99,118 | 51.3 | 35.5 | 13.2 |
| Male | 50,678 | 53.9 | 32.9 | 13.3 |
| Female | 48,229 | 48.7 | 38.1 | 13.2 |
| Nonbinary | 211 | 30.3 | 49.3 | 20.4 |
| EL | 11,929 | 88.5 | 11.3 | 0.2 |
| English only | 50,662 | 46.2 | 37.9 | 15.9 |
| RFEP | 30,917 | 48.7 | 40.0 | 11.3 |
| IFEP | 5,363 | 30.2 | 41.4 | 28.4 |
| ADEL | 207 | 77.8 | 22.2 | 0.0 |
| To be determined | 8 | N/A | N/A | N/A |
| English proficiency unknown | 32 | 59.4 | 37.5 | 3.1 |
| Economically disadvantaged | 61,371 | 60.0 | 32.5 | 7.5 |
| Not economically disadvantaged | 37,747 | 37.2 | 40.2 | 22.6 |
| American Indian or Alaska Native (All) | 449 | 59.0 | 32.7 | 8.2 |
| Asian (All) | 11,099 | 25.4 | 39.5 | 35.1 |
| Native Hawaiian or Other Pacific Islander (All) | 529 | 62.6 | 31.9 | 5.5 |
| Filipino (All) | 2,875 | 29.1 | 48.6 | 22.3 |
| Hispanic or Latino (All) | 55,748 | 60.6 | 32.9 | 6.5 |
| Black or African American (All) | 5,484 | 68.2 | 27.5 | 4.3 |
| White (All) | 18,456 | 39.5 | 40.4 | 20.2 |
| Two or more races (All) | 4,478 | 39.2 | 39.8 | 21.0 |
| Disability | 10,820 | 78.7 | 18.5 | 2.8 |
| No disability | 88,298 | 47.9 | 37.6 | 14.5 |
| Migrant education | 814 | 69.9 | 27.0 | 3.1 |
| Not migrant education | 98,304 | 51.1 | 35.5 | 13.3 |
| Armed forces family member | 1,403 | 51.2 | 35.9 | 12.9 |
| Not armed forces family member | 97,715 | 51.3 | 35.5 | 13.2 |
| Homeless | 3,938 | 68.6 | 27.0 | 4.4 |
| Not homeless | 95,180 | 50.6 | 35.8 | 13.6 |
| Foster youth | 495 | 77.8 | 20.8 | 1.4 |
| Not foster youth | 98,623 | 51.2 | 35.5 | 13.3 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 143 | 53.8 | 33.6 | 12.6 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 306 | 61.4 | 32.4 | 6.2 |
| Asian (Primary ethnicity—not economically disadvantaged) | 6,613 | 17.8 | 39.0 | 43.2 |
| Asian (Primary ethnicity—economically disadvantaged) | 4,486 | 36.5 | 40.3 | 23.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 187 | 51.9 | 39.0 | 9.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 342 | 68.4 | 28.1 | 3.5 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 1,824 | 24.6 | 50.1 | 25.3 |
| Filipino (Primary ethnicity—economically disadvantaged) | 1,051 | 37.0 | 45.9 | 17.1 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 12,368 | 51.3 | 38.0 | 10.7 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 43,380 | 63.3 | 31.4 | 5.3 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 1,623 | 58.3 | 33.8 | 7.9 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 3,861 | 72.4 | 24.9 | 2.7 |
| White (Primary ethnicity—not economically disadvantaged) | 12,272 | 33.5 | 42.3 | 24.2 |
| White (Primary ethnicity—economically disadvantaged) | 6,184 | 51.3 | 36.5 | 12.2 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,717 | 30.6 | 41.7 | 27.6 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,761 | 52.5 | 36.9 | 10.7 |

Table 7.E.12 Percent of Students in Each Performance Level by Demographic Student Group—Life Sciences Domain, High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 454,712 | 46.4 | 38.7 | 15.0 |
| Male | 232,747 | 49.1 | 36.0 | 14.9 |
| Female | 221,020 | 43.6 | 41.4 | 15.0 |
| Nonbinary | 945 | 30.9 | 45.3 | 23.8 |
| EL | 49,860 | 87.0 | 12.7 | 0.3 |
| English only | 242,659 | 41.5 | 40.8 | 17.8 |
| RFEP | 139,846 | 43.5 | 43.6 | 12.9 |
| IFEP | 22,006 | 25.7 | 43.1 | 31.2 |
| ADEL | 248 | 78.2 | 21.8 | 0.0 |
| To be determined | 31 | 77.4 | 12.9 | 9.7 |
| English proficiency unknown | 62 | 50.0 | 40.3 | 9.7 |
| Economically disadvantaged | 283,757 | 54.6 | 36.4 | 9.0 |
| Not economically disadvantaged | 170,955 | 32.6 | 42.4 | 24.9 |
| American Indian or Alaska Native (All) | 1,998 | 54.5 | 34.7 | 10.8 |
| Asian (All) | 45,960 | 21.8 | 41.3 | 36.9 |
| Native Hawaiian or Other Pacific Islander (All) | 2,014 | 55.8 | 35.6 | 8.6 |
| Filipino (All) | 12,041 | 24.8 | 50.1 | 25.1 |
| Hispanic or Latino (All) | 257,977 | 55.3 | 36.6 | 8.1 |
| Black or African American (All) | 21,521 | 63.4 | 30.8 | 5.7 |
| White (All) | 92,569 | 34.5 | 42.9 | 22.5 |
| Two or more races (All) | 20,632 | 35.1 | 41.1 | 23.8 |
| Disability | 49,267 | 78.2 | 19.1 | 2.7 |
| No disability | 405,445 | 42.5 | 41.0 | 16.5 |
| Migrant education | 3,286 | 65.6 | 30.1 | 4.3 |
| Not migrant education | 451,426 | 46.2 | 38.7 | 15.1 |
| Armed forces family member | 6,680 | 46.1 | 40.0 | 13.9 |
| Not armed forces family member | 448,032 | 46.4 | 38.6 | 15.0 |
| Homeless | 14,735 | 65.2 | 29.6 | 5.2 |
| Not homeless | 439,977 | 45.7 | 39.0 | 15.3 |
| Foster youth | 1,761 | 73.0 | 23.7 | 3.4 |
| Not foster youth | 452,951 | 46.2 | 38.7 | 15.0 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 684 | 42.5 | 40.8 | 16.7 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 1,314 | 60.7 | 31.6 | 7.8 |
| Asian (Primary ethnicity—not economically disadvantaged) | 26,227 | 14.9 | 39.8 | 45.3 |
| Asian (Primary ethnicity—economically disadvantaged) | 19,733 | 31.0 | 43.3 | 25.7 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 728 | 46.4 | 41.1 | 12.5 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1,286 | 61.0 | 32.5 | 6.5 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 7,352 | 21.0 | 50.7 | 28.3 |
| Filipino (Primary ethnicity—economically disadvantaged) | 4,689 | 30.9 | 49.1 | 20.0 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 56,798 | 45.8 | 41.1 | 13.1 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 201,179 | 58.0 | 35.4 | 6.6 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 6,161 | 51.6 | 38.0 | 10.4 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 15,360 | 68.2 | 28.0 | 3.8 |
| White (Primary ethnicity—not economically disadvantaged) | 60,556 | 28.5 | 44.2 | 27.3 |
| White (Primary ethnicity—economically disadvantaged) | 32,013 | 46.0 | 40.6 | 13.5 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 12,449 | 26.1 | 43.0 | 30.9 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 8,183 | 48.6 | 38.3 | 13.1 |

Table 7.E.13 Percent of Students in Each Performance Level by Demographic Student Group—Physical Sciences Domain, Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 421,625 | 47.9 | 42.4 | 9.7 |
| Male | 215,155 | 47.6 | 41.9 | 10.5 |
| Female | 206,354 | 48.2 | 42.9 | 8.9 |
| Nonbinary | 116 | 37.1 | 45.7 | 17.2 |
| EL | 80,461 | 80.3 | 19.2 | 0.5 |
| English only | 259,147 | 42.9 | 46.0 | 11.1 |
| RFEP | 55,439 | 33.0 | 55.1 | 11.9 |
| IFEP | 26,446 | 28.9 | 51.5 | 19.6 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 69 | 89.9 | 7.2 | 2.9 |
| English proficiency unknown | 63 | 47.6 | 47.6 | 4.8 |
| Economically disadvantaged | 276,701 | 57.5 | 37.3 | 5.1 |
| Not economically disadvantaged | 144,924 | 29.4 | 52.1 | 18.5 |
| American Indian or Alaska Native (All) | 1,798 | 56.8 | 38.9 | 4.3 |
| Asian (All) | 44,922 | 25.0 | 49.9 | 25.1 |
| Native Hawaiian or Other Pacific Islander (All) | 1,706 | 57.0 | 37.7 | 5.2 |
| Filipino (All) | 9,398 | 27.4 | 56.0 | 16.5 |
| Hispanic or Latino (All) | 234,479 | 58.1 | 37.4 | 4.5 |
| Black or African American (All) | 20,814 | 64.9 | 31.8 | 3.3 |
| White (All) | 84,309 | 33.2 | 51.7 | 15.1 |
| Two or more races (All) | 24,199 | 34.5 | 48.8 | 16.6 |
| Disability | 56,392 | 74.1 | 22.6 | 3.3 |
| No disability | 365,233 | 43.8 | 45.5 | 10.7 |
| Migrant education | 3,430 | 72.0 | 26.3 | 1.7 |
| Not migrant education | 418,195 | 47.7 | 42.5 | 9.8 |
| Armed forces family member | 6,485 | 45.7 | 44.9 | 9.5 |
| Not armed forces family member | 415,140 | 47.9 | 42.4 | 9.7 |
| Homeless | 15,522 | 67.5 | 29.6 | 2.9 |
| Not homeless | 406,103 | 47.1 | 42.9 | 10.0 |
| Foster youth | 1,349 | 74.5 | 23.7 | 1.8 |
| Not foster youth | 420,276 | 47.8 | 42.5 | 9.8 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 464 | 40.9 | 49.1 | 9.9 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 1,334 | 62.3 | 35.4 | 2.3 |
| Asian (Primary ethnicity—not economically disadvantaged) | 25,669 | 16.1 | 51.9 | 32.1 |
| Asian (Primary ethnicity—economically disadvantaged) | 19,253 | 36.9 | 47.3 | 15.8 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 425 | 48.2 | 42.8 | 8.9 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1,281 | 60.0 | 36.1 | 4.0 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 5,310 | 22.6 | 57.2 | 20.3 |
| Filipino (Primary ethnicity—economically disadvantaged) | 4,088 | 33.8 | 54.6 | 11.6 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 45,378 | 43.1 | 47.5 | 9.4 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 189,101 | 61.7 | 35.0 | 3.3 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 4,131 | 46.9 | 45.6 | 7.6 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 16,683 | 69.4 | 28.4 | 2.2 |
| White (Primary ethnicity—not economically disadvantaged) | 50,259 | 24.6 | 56.0 | 19.4 |
| White (Primary ethnicity—economically disadvantaged) | 34,050 | 45.9 | 45.3 | 8.8 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 13,288 | 22.8 | 53.9 | 23.3 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 10,911 | 48.8 | 42.7 | 8.6 |

Table 7.E.14 Percent of Students in Each Performance Level by Demographic Student Group—Physical Sciences Domain, Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 425,511 | 46.9 | 39.0 | 14.1 |
| Male | 218,064 | 47.4 | 37.3 | 15.3 |
| Female | 207,070 | 46.5 | 40.7 | 12.8 |
| Nonbinary | 377 | 29.2 | 52.5 | 18.3 |
| EL | 54,016 | 87.0 | 12.7 | 0.4 |
| English only | 242,072 | 41.9 | 41.5 | 16.6 |
| RFEP | 111,160 | 42.4 | 45.5 | 12.1 |
| IFEP | 18,170 | 22.2 | 44.3 | 33.5 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 57 | 89.5 | 8.8 | 1.8 |
| English proficiency unknown | 36 | 58.3 | 30.6 | 11.1 |
| Economically disadvantaged | 276,684 | 56.5 | 35.9 | 7.6 |
| Not economically disadvantaged | 148,827 | 29.2 | 44.6 | 26.1 |
| American Indian or Alaska Native (All) | 1,791 | 58.5 | 34.2 | 7.3 |
| Asian (All) | 42,397 | 19.9 | 42.2 | 38.0 |
| Native Hawaiian or Other Pacific Islander (All) | 1,842 | 56.2 | 35.6 | 8.3 |
| Filipino (All) | 10,045 | 24.4 | 52.3 | 23.3 |
| Hispanic or Latino (All) | 241,205 | 57.2 | 36.2 | 6.5 |
| Black or African American (All) | 21,254 | 64.8 | 30.3 | 5.0 |
| White (All) | 85,201 | 32.5 | 45.0 | 22.6 |
| Two or more races (All) | 21,776 | 33.3 | 43.0 | 23.7 |
| Disability | 50,212 | 78.3 | 18.9 | 2.8 |
| No disability | 375,299 | 42.7 | 41.7 | 15.6 |
| Migrant education | 3,331 | 68.5 | 27.7 | 3.7 |
| Not migrant education | 422,180 | 46.8 | 39.1 | 14.2 |
| Armed forces family member | 5,972 | 44.4 | 42.1 | 13.5 |
| Not armed forces family member | 419,539 | 47.0 | 38.9 | 14.1 |
| Homeless | 13,656 | 66.8 | 28.8 | 4.4 |
| Not homeless | 411,855 | 46.3 | 39.3 | 14.4 |
| Foster youth | 1,407 | 75.1 | 22.5 | 2.5 |
| Not foster youth | 424,104 | 46.8 | 39.0 | 14.1 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 526 | 40.3 | 47.1 | 12.5 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 1,265 | 66.1 | 28.9 | 5.1 |
| Asian (Primary ethnicity—not economically disadvantaged) | 24,575 | 12.1 | 40.3 | 47.6 |
| Asian (Primary ethnicity—economically disadvantaged) | 17,822 | 30.6 | 44.7 | 24.7 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 516 | 38.0 | 45.5 | 16.5 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1,326 | 63.3 | 31.7 | 5.1 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 5,917 | 20.0 | 52.7 | 27.3 |
| Filipino (Primary ethnicity—economically disadvantaged) | 4,128 | 30.6 | 51.7 | 17.6 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 48,327 | 44.0 | 43.3 | 12.7 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 192,878 | 60.6 | 34.4 | 5.0 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 4,876 | 49.9 | 40.2 | 9.9 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 16,378 | 69.2 | 27.3 | 3.5 |
| White (Primary ethnicity—not economically disadvantaged) | 51,703 | 24.2 | 47.2 | 28.7 |
| White (Primary ethnicity—economically disadvantaged) | 33,498 | 45.3 | 41.5 | 13.2 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 12,387 | 22.3 | 45.3 | 32.4 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 9,389 | 47.8 | 39.9 | 12.3 |

Table 7.E.15 Percent of Students in Each Performance Level by Demographic Student Group—Physical Sciences Domain, Grade Ten

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 24,731 | 53.1 | 37.1 | 9.8 |
| Male | 12,700 | 54.6 | 35.1 | 10.3 |
| Female | 11,967 | 51.6 | 39.2 | 9.2 |
| Nonbinary | 64 | 40.6 | 46.9 | 12.5 |
| EL | 3,032 | 86.5 | 13.0 | 0.4 |
| English only | 13,653 | 50.1 | 39.2 | 10.7 |
| RFEP | 7,020 | 48.1 | 42.5 | 9.4 |
| IFEP | 980 | 26.8 | 45.5 | 27.7 |
| ADEL | 41 | 80.5 | 17.1 | 2.4 |
| To be determined | 0 | N/A | N/A | N/A |
| English proficiency unknown | 5 | N/A | N/A | N/A |
| Economically disadvantaged | 16,556 | 60.0 | 34.3 | 5.6 |
| Not economically disadvantaged | 8,175 | 39.1 | 42.8 | 18.1 |
| American Indian or Alaska Native (All) | 126 | 55.6 | 31.7 | 12.7 |
| Asian (All) | 1,626 | 17.7 | 41.7 | 40.7 |
| Native Hawaiian or Other Pacific Islander (All) | 83 | 59.0 | 36.1 | 4.8 |
| Filipino (All) | 558 | 26.7 | 50.9 | 22.4 |
| Hispanic or Latino (All) | 15,247 | 60.6 | 34.8 | 4.6 |
| Black or African American (All) | 1,442 | 68.7 | 28.6 | 2.7 |
| White (All) | 4,646 | 41.6 | 43.7 | 14.6 |
| Two or more races (All) | 1,003 | 41.4 | 40.3 | 18.3 |
| Disability | 3,053 | 81.2 | 17.4 | 1.4 |
| No disability | 21,678 | 49.2 | 39.9 | 10.9 |
| Migrant education | 233 | 71.2 | 25.8 | 3.0 |
| Not migrant education | 24,498 | 52.9 | 37.2 | 9.8 |
| Armed forces family member | 314 | 51.9 | 37.6 | 10.5 |
| Not armed forces family member | 24,417 | 53.1 | 37.1 | 9.7 |
| Homeless | 753 | 63.1 | 32.7 | 4.2 |
| Not homeless | 23,978 | 52.8 | 37.3 | 9.9 |
| Foster youth | 147 | 71.4 | 25.9 | 2.7 |
| Not foster youth | 24,584 | 53.0 | 37.2 | 9.8 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 39 | 30.8 | 51.3 | 17.9 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 87 | 66.7 | 23.0 | 10.3 |
| Asian (Primary ethnicity—not economically disadvantaged) | 885 | 12.7 | 39.3 | 48.0 |
| Asian (Primary ethnicity—economically disadvantaged) | 741 | 23.6 | 44.5 | 31.8 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 35 | 42.9 | 45.7 | 11.4 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 48 | 70.8 | 29.2 | 0.0 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 351 | 23.4 | 52.4 | 24.2 |
| Filipino (Primary ethnicity—economically disadvantaged) | 207 | 32.4 | 48.3 | 19.3 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 3,202 | 51.0 | 40.8 | 8.2 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 12,045 | 63.2 | 33.2 | 3.7 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 361 | 57.3 | 37.4 | 5.3 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,081 | 72.5 | 25.6 | 1.9 |
| White (Primary ethnicity—not economically disadvantaged) | 2,743 | 34.7 | 46.1 | 19.2 |
| White (Primary ethnicity—economically disadvantaged) | 1,903 | 51.7 | 40.3 | 8.0 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 559 | 32.6 | 41.0 | 26.5 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 444 | 52.5 | 39.4 | 8.1 |

Table 7.E.16 Percent of Students in Each Performance Level by Demographic Student Group—Physical Sciences Domain, Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 330,229 | 47.0 | 38.6 | 14.3 |
| Male | 169,089 | 48.6 | 35.8 | 15.6 |
| Female | 160,471 | 45.4 | 41.5 | 13.0 |
| Nonbinary | 669 | 37.7 | 44.2 | 18.1 |
| EL | 34,802 | 85.5 | 14.0 | 0.4 |
| English only | 178,042 | 42.7 | 40.5 | 16.8 |
| RFEP | 101,702 | 44.6 | 43.3 | 12.2 |
| IFEP | 15,636 | 26.0 | 42.4 | 31.6 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 22 | 68.2 | 27.3 | 4.5 |
| English proficiency unknown | 25 | 56.0 | 28.0 | 16.0 |
| Economically disadvantaged | 205,390 | 55.2 | 36.5 | 8.3 |
| Not economically disadvantaged | 124,839 | 33.5 | 42.2 | 24.3 |
| American Indian or Alaska Native (All) | 1,418 | 56.7 | 34.9 | 8.4 |
| Asian (All) | 33,181 | 21.3 | 40.5 | 38.2 |
| Native Hawaiian or Other Pacific Islander (All) | 1,397 | 53.6 | 36.8 | 9.6 |
| Filipino (All) | 8,596 | 24.8 | 49.6 | 25.6 |
| Hispanic or Latino (All) | 186,601 | 56.1 | 36.7 | 7.2 |
| Black or African American (All) | 14,555 | 63.3 | 31.1 | 5.6 |
| White (All) | 69,352 | 36.2 | 42.7 | 21.0 |
| Two or more races (All) | 15,129 | 36.1 | 40.9 | 22.9 |
| Disability | 35,291 | 79.5 | 18.2 | 2.3 |
| No disability | 294,938 | 43.1 | 41.1 | 15.8 |
| Migrant education | 2,235 | 65.0 | 30.5 | 4.5 |
| Not migrant education | 327,994 | 46.9 | 38.7 | 14.4 |
| Armed forces family member | 4,957 | 47.1 | 39.6 | 13.3 |
| Not armed forces family member | 325,272 | 47.0 | 38.6 | 14.4 |
| Homeless | 10,022 | 66.3 | 29.5 | 4.2 |
| Not homeless | 320,207 | 46.4 | 38.9 | 14.7 |
| Foster youth | 1,116 | 75.0 | 21.9 | 3.1 |
| Not foster youth | 329,113 | 46.9 | 38.7 | 14.4 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 500 | 44.8 | 41.4 | 13.8 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 918 | 63.2 | 31.4 | 5.4 |
| Asian (Primary ethnicity—not economically disadvantaged) | 18,710 | 14.3 | 38.2 | 47.5 |
| Asian (Primary ethnicity—economically disadvantaged) | 14,471 | 30.4 | 43.4 | 26.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 506 | 45.7 | 40.5 | 13.8 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 891 | 58.1 | 34.7 | 7.2 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 5,171 | 21.1 | 49.4 | 29.5 |
| Filipino (Primary ethnicity—economically disadvantaged) | 3,425 | 30.3 | 49.8 | 19.9 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 41,155 | 47.1 | 41.0 | 11.9 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 145,446 | 58.7 | 35.5 | 5.8 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 4,169 | 51.4 | 38.5 | 10.2 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 10,386 | 68.1 | 28.1 | 3.8 |
| White (Primary ethnicity—not economically disadvantaged) | 45,467 | 30.0 | 44.3 | 25.7 |
| White (Primary ethnicity—economically disadvantaged) | 23,885 | 48.0 | 39.8 | 12.2 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 9,161 | 27.2 | 42.9 | 29.9 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 5,968 | 49.8 | 38.0 | 12.2 |

Table 7.E.17 Percent of Students in Each Performance Level by Demographic Student Group—Physical Sciences Domain, Grade Twelve

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 98,844 | 54.1 | 34.1 | 11.8 |
| Male | 50,545 | 55.3 | 31.7 | 13.0 |
| Female | 48,088 | 52.9 | 36.6 | 10.5 |
| Nonbinary | 211 | 37.9 | 48.8 | 13.3 |
| EL | 11,888 | 88.0 | 11.7 | 0.3 |
| English only | 50,534 | 49.7 | 36.4 | 13.9 |
| RFEP | 30,822 | 51.9 | 38.0 | 10.1 |
| IFEP | 5,354 | 31.5 | 41.4 | 27.1 |
| ADEL | 207 | 78.7 | 21.3 | 0.0 |
| To be determined | 7 | N/A | N/A | N/A |
| English proficiency unknown | 32 | 68.8 | 28.1 | 3.1 |
| Economically disadvantaged | 61,193 | 62.6 | 31.3 | 6.1 |
| Not economically disadvantaged | 37,651 | 40.1 | 38.8 | 21.0 |
| American Indian or Alaska Native (All) | 449 | 61.2 | 32.5 | 6.2 |
| Asian (All) | 11,072 | 26.0 | 39.6 | 34.4 |
| Native Hawaiian or Other Pacific Islander (All) | 528 | 59.5 | 34.1 | 6.4 |
| Filipino (All) | 2,863 | 31.4 | 48.1 | 20.5 |
| Hispanic or Latino (All) | 55,580 | 63.6 | 31.3 | 5.1 |
| Black or African American (All) | 5,464 | 71.3 | 25.3 | 3.4 |
| White (All) | 18,421 | 43.0 | 38.8 | 18.1 |
| Two or more races (All) | 4,467 | 43.0 | 39.0 | 18.0 |
| Disability | 10,787 | 80.9 | 16.9 | 2.2 |
| No disability | 88,057 | 50.8 | 36.3 | 13.0 |
| Migrant education | 811 | 72.9 | 25.5 | 1.6 |
| Not migrant education | 98,033 | 53.9 | 34.2 | 11.9 |
| Armed forces family member | 1,399 | 55.3 | 32.8 | 11.9 |
| Not armed forces family member | 97,445 | 54.0 | 34.2 | 11.8 |
| Homeless | 3,923 | 70.6 | 25.8 | 3.7 |
| Not homeless | 94,921 | 53.4 | 34.5 | 12.1 |
| Foster youth | 494 | 80.4 | 18.2 | 1.4 |
| Not foster youth | 98,350 | 53.9 | 34.2 | 11.8 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 143 | 54.5 | 35.0 | 10.5 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 306 | 64.4 | 31.4 | 4.2 |
| Asian (Primary ethnicity—not economically disadvantaged) | 6,596 | 18.6 | 38.2 | 43.3 |
| Asian (Primary ethnicity—economically disadvantaged) | 4,476 | 36.9 | 41.7 | 21.4 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 187 | 50.3 | 38.5 | 11.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 341 | 64.5 | 31.7 | 3.8 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 1,816 | 25.9 | 51.1 | 23.0 |
| Filipino (Primary ethnicity—economically disadvantaged) | 1,047 | 41.1 | 42.8 | 16.1 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 12,336 | 55.1 | 35.7 | 9.2 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 43,244 | 66.0 | 30.0 | 4.0 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 1,615 | 62.0 | 31.3 | 6.6 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 3,849 | 75.1 | 22.8 | 2.0 |
| White (Primary ethnicity—not economically disadvantaged) | 12,247 | 37.1 | 40.8 | 22.1 |
| White (Primary ethnicity—economically disadvantaged) | 6,174 | 54.9 | 34.9 | 10.2 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,711 | 33.6 | 41.9 | 24.5 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,756 | 57.4 | 34.5 | 8.1 |

Table 7.E.18 Percent of Students in Each Performance Level by Demographic Student Group—Physical Sciences Domain, High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Below Standard** | **Near Standard** | **Above Standard** |
| All students | 453,804 | 48.9 | 37.6 | 13.5 |
| Male | 232,334 | 50.4 | 34.9 | 14.7 |
| Female | 220,526 | 47.4 | 40.4 | 12.3 |
| Nonbinary | 944 | 37.9 | 45.4 | 16.6 |
| EL | 49,722 | 86.2 | 13.4 | 0.4 |
| English only | 242,229 | 44.6 | 39.5 | 15.9 |
| RFEP | 139,544 | 46.4 | 42.0 | 11.6 |
| IFEP | 21,970 | 27.4 | 42.3 | 30.3 |
| ADEL | 248 | 79.0 | 20.6 | 0.4 |
| To be determined | 29 | 69.0 | 27.6 | 3.4 |
| English proficiency unknown | 62 | 61.3 | 29.0 | 9.7 |
| Economically disadvantaged | 283,139 | 57.1 | 35.2 | 7.7 |
| Not economically disadvantaged | 170,665 | 35.3 | 41.5 | 23.3 |
| American Indian or Alaska Native (All) | 1,993 | 57.7 | 34.2 | 8.2 |
| Asian (All) | 45,879 | 22.3 | 40.3 | 37.4 |
| Native Hawaiian or Other Pacific Islander (All) | 2,008 | 55.4 | 36.1 | 8.6 |
| Filipino (All) | 12,017 | 26.5 | 49.3 | 24.3 |
| Hispanic or Latino (All) | 257,428 | 58.0 | 35.4 | 6.6 |
| Black or African American (All) | 21,461 | 65.7 | 29.5 | 4.9 |
| White (All) | 92,419 | 37.8 | 42.0 | 20.1 |
| Two or more races (All) | 20,599 | 37.9 | 40.5 | 21.6 |
| Disability | 49,131 | 79.9 | 17.8 | 2.2 |
| No disability | 404,673 | 45.1 | 40.0 | 14.9 |
| Migrant education | 3,279 | 67.4 | 28.9 | 3.7 |
| Not migrant education | 450,525 | 48.8 | 37.6 | 13.6 |
| Armed forces family member | 6,670 | 49.0 | 38.1 | 12.9 |
| Not armed forces family member | 447,134 | 48.9 | 37.6 | 13.5 |
| Homeless | 14,698 | 67.2 | 28.7 | 4.1 |
| Not homeless | 439,106 | 48.3 | 37.9 | 13.9 |
| Foster youth | 1,757 | 76.2 | 21.2 | 2.6 |
| Not foster youth | 452,047 | 48.8 | 37.6 | 13.6 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 682 | 46.0 | 40.6 | 13.3 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 1,311 | 63.7 | 30.8 | 5.5 |
| Asian (Primary ethnicity—not economically disadvantaged) | 26,191 | 15.3 | 38.2 | 46.4 |
| Asian (Primary ethnicity—economically disadvantaged) | 19,688 | 31.6 | 43.1 | 25.4 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 728 | 46.7 | 40.2 | 13.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1,280 | 60.3 | 33.7 | 6.0 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 7,338 | 22.4 | 50.0 | 27.6 |
| Filipino (Primary ethnicity—economically disadvantaged) | 4,679 | 32.8 | 48.2 | 19.0 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 56,693 | 49.0 | 39.8 | 11.1 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 200,735 | 60.6 | 34.2 | 5.3 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 6,145 | 54.5 | 36.5 | 9.0 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 15,316 | 70.2 | 26.6 | 3.2 |
| White (Primary ethnicity—not economically disadvantaged) | 60,457 | 31.6 | 43.7 | 24.7 |
| White (Primary ethnicity—economically disadvantaged) | 31,962 | 49.6 | 38.9 | 11.5 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 12,431 | 28.9 | 42.6 | 28.5 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 8,168 | 51.6 | 37.3 | 11.1 |

## Psychometric Analyses

This chapter describes the psychometric analyses conducted by ETS for the California Science Test (CAST), including classical item analyses, differential item functioning (DIF) analyses, item response theory (IRT) analyses, and response time analyses, as well as analyses to support reliability and validity evidence.

All results presented are for students who participated in the computer-based testing, as only 32 students tested on paper (12 students in grade five, 11 students in grade eight, and 9 students in high school) during the 2023–24 test administration.

### Overview

#### Summary of the Analyses

The following list identifies the analyses that are conducted for a typical CAST administration. Each analysis is described in the subsequent narrative.

1. **Classical Item Analyses—**Classical item analyses for CAST are discussed in section [*8.2 Classical Item Analyses*](#_Classical_Item_Analyses).
2. **DIF Analyses—**DIF analyses for CAST are described in section [*8.3 Differential Item Functioning Analyses*](#_Differential_Item_Functioning).
3. **IRT Analyses—**IRT analyses, including calibration and scaling for CAST, are described in section [*8.4 Item Response Theory Analyses*](#_Item_Response_Theory).
4. **Response Time Analyses—**Response time analyses are described in section [*8.5 Response Time Analyses*](#_Response_Time_Analyses).
5. **Reliability Analyses—**Reliability estimation for CAST is illustrated in section [*8.6 Reliability Analyses*](#_Reliability_Analyses).
6. **Validity Evidence—**Validity evidence related to CAST is discussed in section [*8.7 Validity Evidence*](#_Validity_Evidence).

#### Samples Used for Analyses

In past test administrations, ETS ran two item analyses for CAST: the preliminary item analysis and the final item analysis (FIA). Starting with the 2023–24 test administration, because of the improved process efficiency, FIA can be run in time to allow the results to be available for the data review meeting (DRM). As a result, only the FIA will be conducted for CAST. This will improve the stability of the statistics because more data will be used for the item analysis to inform the DRM; refer to subsection [*11.6.2 Combined Item Analyses*](#_Combined_Item_Analyses) for details.

FIA identifies potentially problematic items for further evaluation and is conducted before the DRM when a sufficient volume of data is collected to obtain stable statistics. In CAST, all student responses to the operational constructed-response (CR) items, and only a sample of student responses from the field test CR items, are scored (refer to subsection [*7.1.1.1.2 Sampling Process for Field Te**st C**onstructed-Response Items*](#_Sampling_Process_for) for details). The FIA included data from both machine-scored and CR items as a single data set for each grade level or the high school grade band (that is, grade ten, eleven, or twelve).

Available student responses that met the inclusion rules were included in the analyses. The inclusion rules used in CAST item analyses and item calibration were as follows:

* Students who logged on to the assessment and answered at least one item were included in the item analysis and item calibration.
* At the item level, items with responses or scores labeled as “omit” were included and treated as “incorrect” for item analyses and calibration. Since CAST requires students to provide a response to move on to the next item, the omit responses are not common. Omit could occur when the student had visited an item but did not provide any response and then exited or paused the test at that item, and the test opportunity subsequently expired. Because the Test Delivery System (TDS) prefetched three items to be displayed, the visited item without responses will be designated as “omit,” and the rest of the prefetched items that were not visited will be designated as “not reached.”
* At the item level, missing responses due to “not reached” or “missing CR scores by design” were excluded from item analyses and calibration. “Not reached” was the result of a student who started the assessment but did not complete it during the testing window.

For score reporting, missing responses for the machine-scorable items due to “omit” were treated as “incorrect.” Not-reached items were not included in the calculation of student scores.

Any field test items flagged during the FIA were sent to the data review committee (refer to subsection [*3.5.4 Data Review Meeting*](#_Data_Review_Meeting) for more details). The California Department of Education (CDE) then made final decisions on the acceptance or rejection of the items on the basis of the data review results. Items that were rejected by the CDE were not included in the IRT calibration process.

#### Test-Taking Rates

Table 8.1 presents the test-taking rates for all grade levels and the high school grade band. Note that test takers are students who were registered for, and logged on to, the assessment. The percentage of registered students who submitted the CAST was 98.18 for grade five, 96.31 for grade eight, and 89.52 for the high school grade band. For high school, 84.69 percent of grade ten students, 92.82 percent of grade eleven students, and 81.04 percent of grade twelve students submitted the CAST. Note that, among the high school grade levels, fewer students tested in grade ten compared to grade eleven and grade twelve.

Table 8.1 CAST Test-Taking Rates of the Full Population

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Grade 5** | **Grade 8** | **HS—Grade 10** | **HS—Grade 11** | **HS—Grade 12** | **HS—All Grades** |
| Number of Registered | 429,092 | 441,145 | 29,080 | 355,094 | 121,646 | 505,820 |
| Number of Started | 422,673 | 428,296 | 24,978 | 332,616 | 99,861 | 457,455 |
| Percent Started | 98.50 | 97.09 | 85.89 | 93.67 | 82.09 | 90.44 |
| Number of Expired | 1,388 | 3,423 | 350 | 3,003 | 1,281 | 4,634 |
| Percent Expired | 0.32 | 0.78 | 1.20 | 0.85 | 1.05 | 0.92 |
| Number of Submitted | 421,285 | 424,873 | 24,628 | 329,613 | 98,580 | 452,821 |
| Percent Submitted | 98.18 | 96.31 | 84.69 | 92.82 | 81.04 | 89.52 |

Table 8.A.1 through table 8.A.6 in [appendix 8.A](#_Alternative_Text_for_17) show the test-taking rates of selected demographic student groups for each assessment. The demographic student groups include economic status, English language fluency, ethnicity, gender, homeless status, migrant status, parent/guardian military status, and disability status.

The grade twelve cohort includes students who tested in grades ten, eleven, or twelve and are enrolled in grade twelve in 2023–24. Table 8.A.7 presents the number of students with valid scores for all grade twelve students within the cohort by demographic student group, regardless of the grade in which they took the CAST.

Table 7.C.1 through table 7.C.6 in [appendix 7.C](#_Appendix_7.C:_Demographic) present the number and percentage of valid test scores for the demographic student groups. Across all grade levels, approximately 50 percent of test takers were male and 50 percent were female. The majority of test takers in all grade levels were English only speakers, economically disadvantaged, Hispanic or Latino, no disability, not eligible for migrant education, not armed forces family member, not homeless, and not foster youth.

### Classical Item Analyses

Classical item analyses are conducted to evaluate the performance of all test items with respect to item difficulty, item-total correlation, and distractor analysis. The associated flagging rules of these statistics are used to identify items that are not performing as expected.

The item analysis has been focused on the computer-based forms for the general population because these are the forms most students take. CAST offered the translations (Spanish stacked–dual language) version of the assessment for students whose primary language was not English and who used dual-language supports in the classroom. (Refer to section [*8.8 Special Considerations for the Spanish-Translation Version*](#_Special_Considerations_for) for additional information on the analyses for this designated support.)

#### Classical Item Difficulty Indices (*p*-value)

Items scored as one (correct) or zero (incorrect) are referred to as dichotomous items. Items scored from zero to some number of points greater than one are called polytomous items.

For dichotomous items, item difficulty is indicated by its *p*-value, which is the proportion of students who answer the item correctly. The range of *p*-values is from 0.00 to 1.00. Items with high *p*-values are easier items; those with low *p*-values are more difficult. Dichotomous items are flagged for review if their *p*-values are above 0.95 (that is, too easy). Two-choice dichotomous single-select items, three-choice dichotomous single-select items, and all other dichotomous items are flagged as too difficult if their *p*-values are below 0.50, 0.30, and 0.20, respectively.

The formula for the *p*-value for a dichotomous item is presented in equation 8.1. *Refer to the* [*Alternative Text for Equation 8.1*](#_Alternative_Text_for_15) *for a description of this equation.*

 (8.1)

where,

*Xij* is the score (0 or 1) received for a given dichotomous item *i* for student *j*, and

*Ji* is the total number of students who were presented with item *i*.

For polytomous items, the difficulty is indicated by either the average item score (AIS) or *p*-‍value. The AIS can range from 0.00 to the maximum total possible points for an item. Desired AIS values for polytomous items generally fall within the range of 20 percent to 80 percent of the maximum obtainable item score; items with values outside this range are flagged for review. To facilitate the interpretation, the AIS values for polytomous items are often expressed as the proportion of the maximum possible score, which are equivalent to the *p-*values for dichotomous items.

For polytomous items, the *p-*value is defined as presented in equation 8.2. *Refer to the* [*Alternative Text for Equation 8.2*](#_Alternative_Text_for_23) *for a description of this equation.*

 (8.2)

where,

*Xij* is the score assigned for a given polytomous item *i* and student *j*,

*Ji* is the total number of students who were presented with item *i*, and

*Mi* is the maximum possible score for item *i*.

#### Item-Total Correlation

An important indicator of item discrimination is the item-total correlation, defined as the correlation between student scores on an individual item and student “total” scores on the assessment.

The item-total correlation statistic describes the relationship between students’ performance on a specific item and students’ performance on the total assessment. It is calculated as the correlation coefficient between the item score and total score—specifically, the polyserial correlation is used as the index of item-total correlation for both polytomous and dichotomous items. Statistically, it is calculated as the correlation between an observed continuous variable and an unobserved continuous variable hypothesized to underlie the variable with ordered categories (Olsson, Drasgow, & Dorans, 1982). The total scale score or the raw score is used as the criterion score for this analysis.

Theoretically, the polyserial correlation ranges from −1.0 (for a perfect negative relationship) to 1.0 (for a perfect positive relationship) and is estimated as presented in equation 8.3. *Refer to the* [*Alternative Text for Equation 8.3*](#_Alternative_Text_for_34) *for a description of this equation.*

 (8.3)

where,

 is an estimated regression coefficient (slope) for predicting the continuous version of an item score onto the continuous version of the total score;

*s2tot* is the variance of the criterion (for example, the students’ total score); and

*stot* is the SD of the criterion.

For a polytomous item, there is a regression for each boundary between item scores, with all regressions for the same item sharing a common slope, *β*. For a polytomous item with *m* possible score values, there are *m−*1 regressions.

Acceptable values for this correlation coefficient are positive and greater than 0.20. A relatively high item-total correlation coefficient value is preferred, as it indicates that higher-performing students tend to perform better on the item than lower-performing students. An item with a negative item-total correlation typically signifies a problem with the item, as that indicates that

* the higher-performing students on the overall assessment tend to respond incorrectly to the item if dichotomous, or are assigned a low score for the item if polytomous; or
* the lower-performing students on the overall assessment are responding correctly to the item if dichotomous, or are assigned a high score for that item if polytomous.

#### Distribution of Item Scores

For polytomous items, examination of the distribution of scores assists in showing how well items performed. If no students were given the highest possible score, the item may not be functioning as expected because the item may be confusing, poorly worded, or just unexpectedly difficult; the scoring rubric may be flawed; or students may not have had an opportunity to learn the content. If the rubric for an item allowed for partial credit but nearly all students received either full credit or partial credit, the rubric should be reviewed for whether the rubric for the partial credit score category should be revised.

Items with a low percentage (that is, less than 3 percent) of students obtaining any score point were flagged for review. Such items may pose problems during IRT calibration. They need to be carefully reviewed and may need to be excluded from the item calibration analyses.

#### Omit Rates

If a student views an item, leaves it unanswered, and then goes on to view and answer another item, the missing response is classified as an “omit.” If the student omits an item—that is, leaves the item unanswered—and does not view additional items, the responses for the successive items are classified as “not seen.”

CAST required students to provide answers to all items on a page before moving on to the next page; therefore, the possibility of an omission would be very small, and it would only happen to one of the last three items that were prefetched by the TDS.

Available student responses that met the inclusion rules were included in the analyses. The inclusion rules used in CAST item analyses and item calibration were as follows:

* Students who logged on to the assessment and answered at least one item were included in the item analysis and item calibration.
* At the item level, items with responses or scores labeled as “omit” were included and treated as “incorrect” for item analyses and calibration. Since CAST requires students to provide a response to move on to the next item, the omit responses are extremely rare and the omit rate at the item level is effectively zero. Omit could occur when the student had visited an item but did not provide any response and then exited or paused the test at that item, and the test opportunity subsequently expired. Because the TDS prefetched three items to be displayed, the visited item without responses will be designated as “omit,” and the rest of the prefetched items that were not visited will be designated as “not reached.”
* At the item level, missing responses due to “not reached” or “missing CR scores by design” were excluded from item analyses and calibration. “Not reached” was the result of a student who started the assessment but did not complete it during the testing window.

For score reporting, missing responses for the machine-scorable items due to “omit” were treated as “incorrect.” Not-reached items were not included in the calculation of student scores.

##### Rates for Dichotomous and Polytomous Items

For both dichotomous and polytomous items, examining the omit rate is useful for identifying potential problems with test features such as testing time and item or test layout. Items with high omit rates are flagged for further investigation by ETS content specialists to ensure that no issues are found with these items. Omit rates for polytomous items tend to be higher than for dichotomous items.

#### Completion Rates

Completion rates are included in table 8.1. The criterion for the minimum number of items, by domain and for overall assessment, is provided in table 7.6.

#### Distractor Analyses

Distractor analyses were conducted on selected-response (SR) items (that is, items that were not constructed response). The statistics for each item included the proportion of students selecting each distractor (incorrect response), computed for the group of all students in the analysis sample, and were also computed separately for the highest-performing 20 percent of students. Items were flagged for review if more high-performing students chose any distractor rather than the key. Such a result indicated that the item may have multiple correct answers or have the wrong key (that is, the item was miskeyed).

For SR items, the distractor-total correlation describes the relationship between selecting a distractor for a specific item and performance on the total assessment. The polyserial correlation was calculated for the distractors, like the item-total correlation previously described, except that the regressions were implemented on the distractors rather than the keys. Items with distractor-total correlations not significantly below zero were flagged for review, as these items may have multiple correct answers, be miskeyed, or have other content issues.

#### Summary of Classical Item Analyses Flagging Criteria

An item was flagged for review if the item analysis yielded any of the following results. One item could have multiple flags if the statistics met the flagging criteria:

* **Difficulty flags** indicated extreme values of the proportion-correct (for dichotomous items) or the proportion of the possible maximum points earned (for polytomous items):
* A-flag: A *p-*value below 0.50 for two-choice dichotomous single-select items, below 0.30 for three-choice dichotomous single-select items, or below 0.20 for all other items
* H-flag: A *p*-value above 0.95 for dichotomous items or above 0.80 for polytomous items
* A **discrimination flag** (R-flag) indicated that the item did not discriminate effectively between high- and low-ability students. Items with a polyserial correlation less than 0.20 were flagged.
* An **omit flag** (O-flag) indicated an omission rate above 5 percent for dichotomous multiple-choice, single-select items or above 15 percent for all other items.
* A **distractor flag** (P-flag) was used for an item with any distractors having a correlation with the criterion score that is either positive, zero, or negative but not significantly below zero.
* A **miskey flag** (D-flag) was used for multiple-choice items when more of the high-ability group of students testing—the top 20 percent of students who are testing on the total assessment—choose any distractor rather than the response keyed as correct.
* An **underrepresented score point flag** (L-flag) was used for any item that had less than 3 percent of the students at any score level.

The ETS Psychometric Analysis & Research staff and AD staff carefully reviewed each of the flagged items during and at the end of the item analyses. All flagged items were also reviewed by California educators at the DRM and then summarized for the CDE with recommendations for subsequent analyses.

#### Classical Item Analyses Results

This subsection provides the summary tables of operational item distributions for the item difficulty and the item discrimination statistics. The overall item difficulty distributions are presented in table 8.2. Across all grade levels and the high school grade band, most items had a *p*-value between 0.2 and 0.8. The percentage of items outside of this range was 4.92 percent for grade five, 0 percent for grade eight, and 5.97 percent for the high school grade band. For high school, the percentage of items outside the desired *p*-value range was 7.46 percent for grades ten and eleven students and 5.97 percent for grade twelve students. Item difficulty distributions by item type are shown in table 8.B.1 in [appendix 8.B](#_Appendix_8.B:_Item); item difficulty distributions by content domain are presented in table 8.B.2.

Table 8.2 Item Difficulty Distributions

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **0 ≤ *p* < 0.2** | **0.2 ≤ *p* < 0.4** | **0.4 ≤ *p* < 0.6** | **0.6 ≤ *p* < 0.8** | **0.8 ≤ *p* ≤ 1.0** | **Total Number of Items** |
| Grade 5 | 2 | 11 | 30 | 17 | 1 | 61 |
| Grade 8 | 0 | 25 | 30 | 6 | 0 | 61 |
| High school—Grade 10 | 5 | 25 | 30 | 7 | 0 | 67 |
| High school—Grade 11 | 4 | 18 | 36 | 8 | 1 | 67 |
| High school—Grade 12 | 4 | 25 | 30 | 8 | 0 | 67 |
| High school—All grades | 4 | 22 | 32 | 9 | 0 | 67 |

Overall item-total correlation distributions are presented in table 8.3. Across all grade levels, with the exception of one item, all items had item-total correlations of 0.2 or higher. No item-total correlations were negative. Most items selected had polyserial correlations higher than 0.30. Item-total correlation distributions by item type are shown in table 8.C.1 in [appendix 8.C](#_Appendix_8.C:_Item-Total); item-total correlation distributions by content domain are presented in table 8.C.2.

Table 8.3 Item-Total Correlation Distributions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **r < 0** | **0 ≤ r < 0.2** | **0.2 ≤ r < 0.3** | **0.3 ≤ r < 0.4** | **0.4 ≤ r < 0.5** | **r ≥ 0.5** | **Total Number of Items** |
| Grade 5 | 0 | 0 | 0 | 3 | 13 | 45 | 61 |
| Grade 8 | 0 | 1 | 3 | 7 | 8 | 42 | 61 |
| High school—Grade 10 | 0 | 0 | 7 | 10 | 15 | 35 | 67 |
| High school—Grade 11 | 0 | 0 | 4 | 8 | 16 | 39 | 67 |
| High school—Grade 12 | 0 | 0 | 5 | 7 | 18 | 37 | 67 |
| High school—All grades | 0 | 0 | 4 | 8 | 16 | 39 | 67 |

### Differential Item Functioning Analyses

DIF is used to evaluate the consistency of individual item performance for students in different demographic student groups who have the same level of domain performance. For example, DIF evaluates whether female and male students matched to have the same test score perform similarly on each item in the assessment.

In examining the DIF between groups, the reference group is often designated as the group that is assumed to have an advantage, while the focal group refers to the group anticipated to possibly be disadvantaged by the assessment.

DIF analyses were conducted for field test items that met the sample size requirements. The sample size requirements for the DIF analyses were 100 in the smaller of either the focal group or the reference group and 400 in the combined focal and reference groups. These sample size requirements are based on standard operating procedures with respect to DIF analyses at ETS.

If an item performs differentially across identifiable student groups—for example, gender or ethnicity—when students are matched on ability, the item may be measuring something else other than the intended construct (that is, possible evidence of bias). It is important, however, to recognize that item performance differences flagged for DIF might be related to actual differences in relevant knowledge or skills between student groups (that is, impact) or statistical Type I error, which might falsely find DIF in an item. As a result, DIF analysis is used mainly as a statistical tool to identify *potential* item bias. Subsequent reviews by content experts and bias and sensitivity experts are required to determine the source and meaning of performance differences.

There are many possible reasons for DIF. The wording of an item, for example, may be such that one group interprets the question differently than the other, or the reading demands of an item are such that, although reading is not being measured (for example, in a mathematics assessment), reading differences between the groups lead to differential outcomes on the item.

DIF analyses were conducted on each assessment for designated comparison groups. Groups are defined on the basis of demographic variables, such as gender, race or ethnicity, and primary disabilities, if the number of students in the group meets the sample size requirements. These comparison groups are specified in table 8.4.

Table 8.4 Student Groups for DIF Comparison

|  |  |  |
| --- | --- | --- |
| **DIF Type** | **Focal Group** | **Reference Group** |
| Gender | Female | Male |
| Ethnicity | American Indian or Alaska Native | White |
| Ethnicity | Asian | White |
| Ethnicity | Black or African American | White |
| Ethnicity | Hispanic or Latino | White |
| English fluency | English learner (EL) | English only |
| Disability | Disability | No disability |
| Economic status | Economically disadvantaged | Not economically disadvantaged |

#### Differential Item Functioning Procedure for Dichotomous Items

The Mantel-Haenszel (MH) DIF (MH-DIF) statistic was calculated for dichotomous items (Mantel & Haenszel, 1959; Holland & Thayer, 1985). For this method, students are classified into relevant student groups of interest (for example, gender or ethnicity). Students at each total score level in the focal group (for example, females) are compared with students at each total score level in the reference group (for example, males). The common odds ratio—that is, the proportion of correct response over the proportion of incorrect response—is estimated across all levels of matched student ability using the formula in equation 8.4 (Dorans & Holland, 1993). The resulting estimate is interpreted as the relative probability of success on a particular item for members of two groups when matched on ability. *Refer to the* [*Alternative Text for Equation 8.4*](#_Alternative_Text_for_42) *for a description of this equation.*

 (8.4)

where,

*M* is the highest score category of the criterion score (total raw score),

*m* indexes the score categories,

*Rrm* is the number of students in the reference group at score level *m* who answer the item correctly,

*Wfm* is the number of students in the focal group at score level *m* who answer the item incorrectly,

*Ntm* is the total number of students at score level *m*,

*Rfm* is the number of students in the focal group at score level *m* who answer the item correctly, and

*Wrm* is the number of students in the reference group at score level *m* who answer the item incorrectly.

To facilitate the interpretation of MH results, the common odds ratio is frequently transformed onto the delta scale using equation 8.5 (Holland & Thayer, 1985). *Refer to the* [*Alternative Text for Equation 8.5*](#_Alternative_Text_for_43) *for a description of this equation.*

 (8.5)

Positive values indicate DIF in favor of the focal group (that is, positive DIF items are differentially easier for the focal group), whereas negative values indicate DIF in favor of the reference group (that is, negative DIF items are differentially easier for the reference group).

#### Differential Item Functioning Procedure for Polytomous Items

The standardization DIF (Dorans & Schmitt, 1993; Zwick, Thayer, & Mazzeo, 1997; Dorans, 2013) in conjunction with the Mantel chi-square statistic (Mantel, 1963; Mantel & Haenszel, 1959) is calculated for polytomous items. The standardized mean difference (SMD) compares the item means of the two groups after adjusting for differences in the distribution of students across all items and is calculated using equation 8.6. *Refer to the* [*Alternative Text for Equation 8.6*](#_Alternative_Text_for_44) *for a description of this equation.*

 (8.6)

where,

*M* is the highest score category of the criterion score (total raw score),

*Nfm* is the number of students in the focal group at score level *m*,

*Erm* is the expected item score for the reference group at score level *m*,

*Efm* is the expected item score for the focal group at score level *m*, and

*Dm* is the difference in the distribution of students at score level *m*.

These statistics are indicators of the degree to which members of one group perform better or worse than expected on each polytomous item.

A positive SMDvalue means that, conditional on the criterion score, the focal group has a higher mean item score than the reference group (that is, the item is differentially easier for the focal group). In contrast, a negative SMD value means that, conditional upon the criterion score, the focal group has a lower mean item score than the reference group (that is, the item is differentially harder for the focal group).

#### Classification

Based on the DIF statistic values and significance tests, items are classified into three categories and assigned values of A, B, or C (Holland & Wainer, 1993). Category A items contain negligible DIF, Category B items exhibit slight to moderate DIF, and Category C items possess moderate to large DIF values. Items flagged as C-level DIF (including C+ and C−) will be brought to the DRM for educator review.

The flagging criteria for dichotomous items are presented in table 8.5; the flagging criteria for polytomous items are provided in table 8.6. The determination of all significant differences is based on *p*-value < 0.05.

Table 8.5 DIF Categories for Dichotomous Items

|  |  |
| --- | --- |
| **DIF Category** | **Criteria** |
| A (negligible) | * Absolute value of MH D-DIF is less than one or is not significantly different from zero. * Positive values are classified as “A+” and negative values as “A−.” |
| B (moderate) | * Absolute value of MH D-DIF is significantly different from zero but not from one and is at least one; or absolute value of MH D-DIF is significantly different from one but is less than 1.5. * Positive values are classified as “B+” and negative values as “B−.” |
| C (large) | * Absolute value of MH D-DIF is at least 1.5 and is significantly different from one. * Positive values are classified as “C+” and negative values as “C−.” |

In table 8.6, SMD is standardized mean difference and SD is total group standard deviation of the item score.

Table 8.6 DIF Categories for Polytomous Items

|  |  |
| --- | --- |
| **DIF Category** | **Criteria** |
| A (negligible) | Mantel chi-square *p-*value≥ 0.05 or |SMD/SD| ≤ 0.17 |
| B (moderate) | Mantel chi-square *p-*value *<* 0.05 and 0.17 < |SMD/SD| ≤ 0.25 |
| C (large) | Mantel chi-square *p-*value *<* 0.05 and |SMD*/*SD| > 0.25 |

#### Differential Item Functioning Analysis Results

Summarized DIF results are presented in table 8.7 through table 8.9 for grade five, grade eight, and high school, respectively.

Two items in grade five were identified as having C+ DIF. One item in grade eight and one item in high school were identified as having C− DIF. In total, four items were flagged for C-‍level DIF. Prior to placement on an operational test form, the items that show C-level DIF—which is a *statistical* flag—are first reviewed by the DIF review panel. The DIF panel determines whether the items are biased or unfair. Items that are biased or unfair are deactivated for further use.

Assessment developers were instructed to avoid selecting C-level items considered unbiased by the DIF review panel for future test forms unless their inclusion is deemed essential to meeting test-content specifications and is approved by the CDE.

Table 8.7 Number of Items by DIF Category for Grade Five

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Focal Group–Reference Group** | **DIF Category A** | **DIF Category B−** | **DIF Category B+** | **DIF Category C−** | **DIF Category C+** | **Insufficient Counts** |
| Female–Male | 61 | 0 | 0 | 0 | 0 | 0 |
| Asian–White | 59 | 1 | 1 | 0 | 0 | 0 |
| Black–White | 61 | 0 | 0 | 0 | 0 | 0 |
| Filipino–White | 58 | 1 | 0 | 0 | 2 | 0 |
| Hispanic–White | 61 | 0 | 0 | 0 | 0 | 0 |
| American Indian or Alaska Native–White | 61 | 0 | 0 | 0 | 0 | 0 |
| Native Hawaiian–White | 60 | 0 | 1 | 0 | 0 | 0 |
| Two or more races–White | 61 | 0 | 0 | 0 | 0 | 0 |
| EL–English only | 61 | 0 | 0 | 0 | 0 | 0 |
| Disability–No disability | 60 | 1 | 0 | 0 | 0 | 0 |
| Economically disadvantaged–Not economically disadvantaged | 61 | 0 | 0 | 0 | 0 | 0 |

Table 8.8 Number of Items by DIF Category for Grade Eight

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Focal Group–Reference Group** | **DIF Category A** | **DIF Category B−** | **DIF Category B+** | **DIF Category C−** | **DIF Category C+** | **Insufficient Counts** |
| Female–Male | 59 | 1 | 0 | 1 | 0 | 0 |
| Asian–White | 60 | 0 | 1 | 0 | 0 | 0 |
| Black–White | 61 | 0 | 0 | 0 | 0 | 0 |
| Filipino–White | 57 | 1 | 3 | 0 | 0 | 0 |
| Hispanic–White | 61 | 0 | 0 | 0 | 0 | 0 |
| American Indian or Alaska Native–White | 61 | 0 | 0 | 0 | 0 | 0 |
| Native Hawaiian–White | 61 | 0 | 0 | 0 | 0 | 0 |
| Two or more races–White | 61 | 0 | 0 | 0 | 0 | 0 |
| EL–English only | 60 | 1 | 0 | 0 | 0 | 0 |
| Disability–No disability | 61 | 0 | 0 | 0 | 0 | 0 |
| Economically disadvantaged–Not economically disadvantaged | 61 | 0 | 0 | 0 | 0 | 0 |

Table 8.9 Number of Items by DIF Category for High School

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Focal Group–Reference Group** | **DIF Category A** | **DIF Category B−** | **DIF Category B+** | **DIF Category C−** | **DIF Category C+** | **Insufficient Counts** |
| Female–Male | 62 | 0 | 4 | 1 | 0 | 0 |
| Asian–White | 65 | 0 | 2 | 0 | 0 | 0 |
| Black–White | 66 | 1 | 0 | 0 | 0 | 0 |
| Filipino–White | 62 | 1 | 4 | 0 | 0 | 0 |
| Hispanic–White | 66 | 0 | 1 | 0 | 0 | 0 |
| American Indian or Alaska Native–White | 67 | 0 | 0 | 0 | 0 | 0 |
| Native Hawaiian–White | 67 | 0 | 0 | 0 | 0 | 0 |
| Two or more races–White | 67 | 0 | 0 | 0 | 0 | 0 |
| EL–English only | 64 | 2 | 1 | 0 | 0 | 0 |
| Disability–No disability | 67 | 0 | 0 | 0 | 0 | 0 |
| Economically disadvantaged–Not economically disadvantaged | 67 | 0 | 0 | 0 | 0 | 0 |
| Grade ten–Grade eleven | 67 | 0 | 0 | 0 | 0 | 0 |
| Grade twelve–Grade eleven | 67 | 0 | 0 | 0 | 0 | 0 |

### Item Response Theory Analyses

IRT is a family of mathematical models that characterizes the probability of a given response as a function of a student’s true ability and one or more features of the items, such as its difficulty or discrimination. IRT can be used to calibrate items, link item parameter estimates, scale or equate test scores across different forms or test administrations, evaluate item performance, build an item bank, and assemble test forms.

This section describes how IRT models were used to calibrate and link field test items onto the base IRT scale established during the 2018–19 administration. Only items that were not rejected by both the data review committees and the CDE were included in the calibration process.

#### Item Response Theory Model

The two-parameter logistic IRT model was used to calibrate the dichotomous items (that is, items worth 1 point) and the generalized partial credit model (GPCM) (Muraki, 1992) was used to calibrate the polytomous items (that is, items worth more than 1 point). FlexMIRT® (Cai, 2017), a multilevel and multiple-group IRT software package, was used for the calibration.

The mathematical form of the GPCM is presented in equation 8.7. *Refer to the* [*Alternative Text for Equation 8.7*](#_Alternative_Text_for_45) *for a description of this equation.*

 (8.7)

where,

 is the probability of student with proficiency  obtaining score *h* on item *i*,

*Mi* is the maximum possible score points for item *i*,

*ai* is the discrimination parameter,

*bi* is the location parameter for item *i*,

*div* is the category parameter for item *i* on item score *v*,

*D* is a scaling constant of 1.7,

*c* indexes the item score, and

*v* indexes the nonzero item score.

When *Mi* = 1, equation 8.7 becomes an expression of the two-parameter logistic model for dichotomous items.

#### Data Preparation

Prior to IRT calibration analyses, ETS psychometricians reviewed the results of the classical item analyses to decide whether any items were of poor quality and needed to be removed from calibration. The results also were reviewed by ETS content experts and the CDE. The decision whether to remove items from calibration was made in consultation with the CDE.

For IRT calibration, scored item response data was used to create the IRT analysis input data files for each grade level. The IRT analysis input data file was a full matrix containing item-level scores for students who answered at least one item.

Similar to the classical item analyses, “omit” items were treated as incorrect and “not presented” items were treated as blank.

#### Equating

Equating is a procedure where test scores, from different test forms assembled on the basis of the same specifications, are placed onto a reference scale so that scores from different test administrations are comparable. There are two approaches to equate the test forms: preequating and postequating.

A preequating design allows for conversion tables that describe the relationship between raw scores and scale scores, or theta scores and scale scores, to be established prior to the current test administration using data from prior administrations. Preequating relies on having a well-calibrated item bank, robust embedded field-testing processes, and stability in item performance over time.

A postequating design uses the data from the current administration to establish the raw-to-scale-score relationship for the current administration’s form.

Both preequating and postequating involve a common‑item nonequivalent groups design (Kolen & Brennan, 2004).

For all assessments, regardless of whether they are preequated or postequated, IRT calibration and linking were conducted to put the field test item parameters onto the base IRT scale.

CAST was preequated to the base scale established in the 2018–19 administration. All operational items have parameters that were already linked to the base scale through the field testing from administrations prior to the current administration. For detailed information on the method to establish the raw-score-to-scale-score relationship, refer to subsection [*8.4.6* *Scaling the Scores*](#_Scaling_the_Scores)*.*

#### Calibration and Linking for the Field Test Items

After each administration, the field test items will be calibrated and linked to the base scale.

##### Calibration

The calibration will be conducted using a sparse matrix combining all operational items and field test items from all versions of the forms within a grade level or grade span.

FlexMIRT (Cai, 2017), a multilevel and multiple-group IRT software package for item analysis and test scoring, was used for item calibration analysis. This software can fit a variety of IRT models to both single-level and multilevel data that are dichotomous, polytomous, or both, and was chosen for its superior flexibility among IRT software programs.

The evaluation of the calibration results includes the following steps:

1. Reviewing the item parameter estimates to examine whether these estimates were reasonable
   1. At the form level, the summary statistics for the *b*-parameter estimates (location difficulty) and *d*-parameter estimates (step difficulty) were examined, including the mean, SD, median, minimum, maximum, and goodness-of-fit.
   2. At the item level, statistics of individual items were examined, including item difficulty estimates, model-fit statistics, and the IRT-based item parameters.
2. Flagging items that did not perform as expected (All flagged items were discussed thoroughly with the CDE to decide whether those items should be removed from calibration or whether the scoring categories need to be collapsed.)

The calibration process was paralleled by two ETS psychometricians to ensure quality and accuracy of results. Specifically, two psychometricians independently created flexMIRT control files and ran the same input data files and then compared the calibration results. Any differences in the output were investigated. Refer to section [*9.6 Quality Control of Psychometric Processes*](#_Quality_Control_of_1)for more details of this procedure.

##### Linking

The item parameters obtained through the calibration are on a different scale and will be linked to the baseline scale using all operational items as anchors.

The Stocking-Lord (1983) method will be applied for the linking. The software STUIRT (Kim & Kolen, 2004) is employed to find the Stocking-Lord linking constants for the common items. The stability of the anchor items will be evaluated by means of item response functions (IRFs) and a d-square measure. The d-square measure is the weighted squared deviation of the IRFs across the range of proficiency (that is, theta or *θ*) and under a hypothetical normal distribution for *θ* between the item parameter estimates from the calibration and the equated values. For a given item , the measure d squared sub i or “d-squared” is defined as follows. *Refer to the* [*Alternative Text for Equation 8.8*](#_Alternative_Text_for_46) *for a description of this equation.*

 (8.8)

where,

*i* indexes common or anchor items,

*k* indexes quadrature points for *θ*,

*si,new* is the expected item score for item *i* based on transformed parameter estimates of operational items from the current administration,

while *si,bank* is the expected item score for item *i* based on the item bank parameter estimates, and

*g(θk)* are weights for each of the *k* quadrature points *θk*.

The magnitude of d-square in conjunction with plots of these curves are used to identify items for removal from the anchor set. The frequently used minimum d-squared criteria for anchor item removal are 0.1252 for dichotomous items and 0.22 for polytomous items. For anchor items with d-squared results that are at the borderline of these thresholds, the evaluation will be made on an individual basis.

#### Parameter Estimates

The range of *a*-parameter estimates was between 0.14 and 1.55 across the two grade levels and the high school grade band. The means of *a*-parameter estimates were 0.68 for grade five, 0.66 for grade eight, and 0.67 for high school. In addition, the summaries of the IRT *a-‍*parameter estimates for each grade-level assessment are presented in [appendix 8.D](#_Appendix_8.D:_Item), in table 8.D.1 through table 8.D.3 by item type; and table 8.D.4 through table 8.D.6 by content domain for grades five and eight and high school, respectively.

The overall summary of the IRT *a*-parameter estimates is shown in table 8.10. The number of items in each of the *a*-parameter intervals is shown for grades five and eight and for high school. The summary statistics, such as the minimum, maximum, mean, and SD values of the item parameters are also presented.

Table 8.10 Item Discrimination Parameter Distribution by Grade Level or Grade Band

|  |  |  |  |
| --- | --- | --- | --- |
| **IRT-a Range** | **Grade 5** | **Grade 8** | **High School** |
| a < 0 | 0 | 0 | 0 |
| 0 ≤ a < 0.2 | 0 | 2 | 1 |
| 0.2 ≤ a < 0.4 | 6 | 10 | 11 |
| 0.4 ≤ a < 0.6 | 19 | 16 | 23 |
| 0.6 ≤ a < 0.8 | 15 | 12 | 12 |
| 0.8 ≤ a < 1.0 | 16 | 13 | 10 |
| 1.0 ≤ a < 1.2 | 3 | 7 | 7 |
| 1.2 ≤ a < 1.4 | 2 | 1 | 1 |
| 1.4 ≤ a < 1.6 | 0 | 0 | 2 |
| 1.6 ≤ a < 1.8 | 0 | 0 | 0 |
| 1.8 ≤ a < 2.0 | 0 | 0 | 0 |
| a ≥ 2.0 | 0 | 0 | 0 |
| Minimum | 0.30 | 0.14 | 0.20 |
| Maximum | 1.27 | 1.30 | 1.55 |
| Mean | 0.68 | 0.66 | 0.67 |
| SD | 0.23 | 0.27 | 0.31 |
| **Number of Items:** | **61** | **61** | **67** |

Similar information for the IRT *b*-parameter estimates is shown in table 8.11 for the number of items in each of the *b*-parameter intervals and the summary statistics such as the minimum, maximum, mean, and SD values for each grade level and the high school grade band. The means of *b*-parameter estimates were −0.05, 0.41, and 0.46 for grades five and eight and for the high school grade band, respectively, indicating that the mean item difficulty level increased slightly as the grade level increased. All items had *b-*parameter estimates within the acceptable range of −4 to +4.

The summaries of *b*-parameter estimates, separated by item type, are shown in [appendix 8.E](#_Appendix_8.E:_Item) in table 8.E.1 through table 8.E.3 and by content domain in table 8.E.4 through table 8.E.6 for grades five and eight, and for high school, respectively.

Table 8.11 Item Difficulty Parameter Distribution by Grade Level or Grade Band

|  |  |  |  |
| --- | --- | --- | --- |
| **IRT-*b* Range** | **Grade 5** | **Grade 8** | **High School** |
| b < −3.5 | 0 | 0 | 0 |
| −3.5 ≤ b < −3.0 | 0 | 0 | 0 |
| −3.0 ≤ b < −2.5 | 0 | 0 | 0 |
| −2.5 ≤ b < −2.0 | 0 | 0 | 0 |
| −2.0 ≤ b < −1.5 | 1 | 0 | 1 |
| −1.5 ≤ b < −1.0 | 5 | 0 | 2 |
| −1.0 ≤ b < −0.5 | 10 | 4 | 5 |
| −0.5 ≤ b < 0 | 17 | 14 | 15 |
| 0 ≤ b < 0.5 | 14 | 18 | 15 |
| 0.5 ≤ b < 1.0 | 9 | 11 | 16 |
| 1.0 ≤ b < 1.5 | 3 | 10 | 4 |
| 1.5 ≤ b < 2.0 | 2 | 4 | 5 |
| 2.0 ≤ b < 2.5 | 0 | 0 | 1 |
| 2.5 ≤ b < 3.0 | 0 | 0 | 1 |
| 3.0 ≤ b < 3.5 | 0 | 0 | 2 |
| b ≥ 3.5 | 0 | 0 | 0 |
| Minimum | −1.63 | −0.88 | −1.92 |
| Maximum | 1.98 | 1.90 | 3.16 |
| Mean | −0.05 | 0.41 | 0.46 |
| SD | 0.73 | 0.70 | 0.94 |
| **Number of Items:** | **61** | **61** | **67** |

#### Scaling the Scores

For CAST, a four-step procedure is used to obtain scale scores. First, for each assessment, a test characteristic curve (TCC) is computed by summing the IRFs across all items. A theta score is then obtained by “inverting” this TCC for every raw score on the assessment. This theta score is then transformed to a number correct score on a reference form scale. Finally, a linear transformation is used to transform the number correct score to the final scale score. The details on how the scales were established are described in subsection[*8.4.6.2 Transformation from Theta Scores to Scale Scores*](#_Transformation_from_Theta_1).

##### Inverse Test Characteristic Curve Procedure

After all the item difficulty estimates are transformed to the reference scale, students’ overall ability estimates can be derived from the input data file that was described in subsection [*8.4.2 Data Preparation*](#_Data_Preparation)*,* through the IRT inverse TCC method (Stocking, 1996). This method transforms the sum of the student’s item scores into an ability estimate. That estimate is the ability value that makes the sum of the expected scores on the items administered to the student equal to the sum of the scores that the student actually received on those items.

The TCC expresses the expected total score on a set of items as a function of the student’s ability, which is shown in equation 8.9. *Refer to the* [*Alternative Text for Equation 8.9*](#_Alternative_Text_for_47) *for a description of this equation.*

 (8.9)

where,

*i* indexes items,

*j* indexes students,

*I* is the total number of items in the assessment,

*Mi* is the maximum possible score for item *i*,

*h* indexes the value for each score category for item *i*,

is the probability of examinee with ability  obtaining score *h* on item *i* in equation 8.7, and

*ξ(θ)* is the corresponding expected total score.

##### Transformation from Theta Scores to Scale Scores

After the raw scores are converted to theta scores using the inverse TCC method described in the previous subsection, the theta scores are converted to number-correct scores on a reference form using [equation 7.1](#EQ7_1), as described in subsection [*7.2.2* *Scale Scores for the Total Assessment*](#_Scale_Scores_for)*.* This number-correct score on the reference form can then transformed to a scale score using a linear transformation, as presented in equation 8.10. *Refer to the* [*Alternative Text for Equation 8.10*](#_Alternative_Text_for_48) *for a description of this equation.*

 (8.10)

where,

*SSj* is the scale score for student *j*,

 is the adjusted[[9]](#footnote-10) number-right (NR) score for student *j* on the base form given the estimated theta score , and

*A\** and *B\** are the linear transformation constants to transform the expected raw score on a reference form to the scale score.

Because different-length forms are administered across students, and some students did not complete the assessment, a proportional adjustment on NR is used to provide an equitable score to all students. The adjustment for incomplete test takers is proportional to the fraction of the assessment completed.

The transformation constants *A\** and *B\** in equation 8.10 are derived by mapping the lowest obtainable adjusted NR (LOANR) score and the highest obtainable adjusted NR (HOANR) score to the lowest obtainable scale score (LOSS) plus one and the highest obtainable scale score (HOSS) (that is, the A and B parameters are derived by solving equation 8.11). *Refer to the* [*Alternative Text for Equation 8.11*](#_Alternative_Text_for_35) *for a description of this equation.*

 (8.11)

The use of LOSS+1 instead of LOSS is to differentiate students who answered less than 10 items from those who answered at least 10 items but did not get any items correct. The former will receive LOSS and the latter will receive LOSS+1 as the reported scores. The LOSS used in equation 8.11 is 150 for grade five, 350 for grade eight, and 550 for high school. The HOSS is 250 for grade five, 450 for grade eight, and 650 for high school.

The LOANR and HOANR are the lowest obtainable NR (LONR) score and the highest obtainable NR (HONR) score adjusted by the proportion of items answered. The LONR and HONR are determined by the lowest obtainable theta (LOT) and the highest obtainable theta (HOT) scores through the test characteristic function of the base form, which comprises 100 Rasch items with a difficulty of 0, as presented in equation 8.12 (*Refer to the* [*Alternative Text for Equation 8.12*](#_Alternative_Text_for_36) *for a description of this equation.)*

 (8.12)

and equation 8.13 *(Refer to the* [*Alternative Text for Equation 8.13*](#_Alternative_Text_for_7) *for a description of this equation.)*

 (8.13)

The scaling constants *A\** and *B\** derived using the preceding method for grade five, grade eight, and high school are provided in table 7.5 in subsection [*7.2.2* *Scale Scores for the Total Assessment*](#_Scale_Scores_for)*.*

### Response Time Analyses

CAST includes three segments: Segment A (operational discrete items), Segment B (operational performance tasks [PTs]), and Segment C (field test items). Each student received two blocks in Segment A, three PTs in Segment B, and either one PT or one block of discrete items in Segment C. CAST is an untimed assessment.[[10]](#footnote-11)

The estimated time, according to the test design, for students to complete the assessment was 60 minutes for Segment A, 40 minutes for Segment B, and 20 minutes for Segment C. The time it took students to complete an assessment was recorded and analyzed.

Response time analyses were based on students who logged on to the assessment and whose total testing time at the test level did not equal zero. According to the test design, half of the students received a PT block and the other half of the students received a discrete item block in Segment C. Therefore, response time analyses for Segment C were conducted separately for the PT block and the discrete item block.

Because the testing time for a discrete block may be different from that of a PT block, the testing time for the total assessment in table 8.12 was broken down for students who received a PT in Segment C (that is, two Segment A blocks + three PTs + one field test PT) and those who received a discrete block in Segment C (that is, two Segment A blocks + three PTs + one field test discrete block). The unit of testing time is minutes.

Some cases of extremely long testing time may be attributed to students with special needs taking longer to complete the assessment, or the assessment’s not being closed down properly. Therefore, mean testing times may be misleading. The medians (fiftieth percentile) are used to interpret the results because medians are less impacted by the extreme values and are, therefore, more indicative of the central tendency of the data compared to the mean for evaluating testing time information.

The median total testing time was similar for students receiving a form with a discrete Segment C block and students receiving a form with a PT Segment C block. The median total testing time decreased as the grade level increased. Note that the criterion for students to be included in table 8.12 is that none of their items are designated as “not seen” items.

Table 8.12 Testing Time (in Minutes) for the Total Assessment

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Segment** | **N** | **Mean** | **SD** | **Min** | **Max** | **1st Percentile** | **10th Percentile** | **25th Percentile** | **50th Percentile** | **75th Percentile** | **90th Percentile** | **99th Percentile** |
| Grade 5 | 2A Blocks + 3 PTs + 1 Field Test Discrete Block | 210,747 | 117.0 | 67.4 | 1.1 | 1150.9 | 20.8 | 52.9 | 73.3 | 102.1 | 143.4 | 197.5 | 357.8 |
| Grade 5 | 2A Blocks + 3 PTs + 1 Field Test PT | 210,570 | 118.0 | 67.9 | 2.4 | 1174.6 | 20.3 | 53.0 | 73.6 | 103.1 | 145.3 | 199.5 | 359.7 |
| Grade 8 | 2A Blocks + 3 PTs + 1 Field Test Discrete Block | 212,468 | 96.6 | 53.5 | 1.5 | 1199.0 | 12.4 | 42.8 | 62.4 | 86.9 | 119.2 | 160.2 | 280.2 |
| Grade 8 | 2A Blocks + 3 PTs + 1 Field Test PT | 212,448 | 96.1 | 53.6 | 2.5 | 967.1 | 12.2 | 42.2 | 61.8 | 86.4 | 118.8 | 159.5 | 281.7 |
| High school—Grade 10 | 2A Blocks + 3 PTs + 1 Field Test Discrete Block | 12,257 | 77.4 | 38.4 | 0.4 | 854.1 | 10.5 | 35.5 | 53.0 | 73.0 | 95.3 | 121.5 | 195.9 |
| High school—Grade 10 | 2A Blocks + 3 PTs + 1 Field Test PT | 12,387 | 76.8 | 38.0 | 0.9 | 480.6 | 10.1 | 34.2 | 52.1 | 72.5 | 95.2 | 122.5 | 196.4 |
| High school—Grade 11 | 2A Blocks + 3 PTs + 1 Field Test Discrete Block | 164,588 | 73.2 | 41.4 | 0.5 | 811.5 | 7.7 | 27.9 | 47.2 | 68.1 | 91.5 | 121.0 | 209.2 |
| High school—Grade 11 | 2A Blocks + 3 PTs + 1 Field Test PT | 165,082 | 73.1 | 41.3 | 1.6 | 1040.6 | 7.4 | 27.8 | 47.2 | 68.2 | 91.0 | 120.2 | 208.3 |
| High school—Grade 12 | 2A Blocks + 3 PTs + 1 Field Test Discrete Block | 49,364 | 61.1 | 35.8 | 1.8 | 723.5 | 6.3 | 20.7 | 37.7 | 56.9 | 77.7 | 102.8 | 176.1 |
| High school—Grade 12 | 2A Blocks + 3 PTs + 1 Field Test PT | 49,268 | 61.0 | 35.8 | 2.2 | 914.8 | 6.1 | 20.5 | 37.4 | 56.9 | 77.7 | 102.8 | 176.0 |
| High school—All grades | 2A Blocks + 3 PTs + 1 Field Test Discrete Block | 226,209 | 70.8 | 40.4 | 0.4 | 854.1 | 7.3 | 26.3 | 45.0 | 65.9 | 89.0 | 117.5 | 203.0 |
| High school—All grades | 2A Blocks + 3 PTs + 1 Field Test PT | 226,737 | 70.6 | 40.3 | 0.9 | 1040.6 | 7.0 | 26.1 | 45.0 | 65.9 | 88.7 | 117.1 | 202.0 |

Table 8.F.1 in [appendix 8.F](#_Appendix_8.F:_Response) shows the testing time by segment. The testing time for an operational discrete block was longer than that for an operational PT block for all the percentiles considered because there are far more items in the discrete block.

Table 8.F.2 shows the testing time per item for four item types: multiple choice (MC), CR, technology enhanced, and composite. Because testing time was recorded at the page level, PT items that were on a page with multiple items—around 30 to 50 percent of the items—were excluded from the analysis for table 8.F.2. The testing time for a CR item was the longest for each percentile across all grade levels and the high school grade band.

Table 8.F.3 shows the testing time for the operational items (segments A and B) by content domain. Segment C was excluded from the analysis because students receive PTs from different domains in Segment C. The criterion for students to be included in this table is that none of their items in segments A and B are designated as “not seen” items.

The median testing time for operational items varied slightly across the three domains. For grade five, students spent less time in the PS domain, while the median testing time for the ESS and LS domains were similar. Grade eight students spent more time in the LS domain and less time in the ESS domain. For students in the high school grade band, the median testing time was slightly longer for the ESS domain and slightly shorter for the PS domain.

### Reliability Analyses

The reliability for a particular group of students’ test scores is the extent to which the scores would remain consistent if those same students were retested with a parallel version of the same assessment. There are many definitions of reliability (Haertel, 2006) that have their genesis in classical test theory and a variety of methods that can be used to estimate reliability.

The general concept of reliability concerns the extent to which the test scores measure *a particular construct* consistently. The variance in the distribution of test scores—essentially, the observed differences among individuals—is partly due to differences that are consistent and partly due to differences that are not consistent. The measure of variation associated with the first kind of differences—consistent differences—is called “true variance”; this would include actual differences in students’ knowledge. The measure of variation associated with the remaining differences—those that operate essentially at random—is called “error variance.” Error variance includes a variety of underlying differences such as selections of test content, which may cause a student’s test score to be slightly higher in one evaluation and slightly lower in another. Reliability is the proportion of total variance that is due to true variance. The standard error of measurement (SEM) is a statistic that characterizes the error variance.

Reliability coefficients range from zero to one. The higher the reliability coefficient for a set of scores, the more likely individuals are to obtain very similar scores upon repeated testing occasions, if the students do not change in their level of the knowledge or skills measured by the assessment.

#### Sample for Reliability Analyses

The reliability analyses performed for CAST require that the sample be screened beyond the requirements listed in subsection [*8.1.2 Samples Used for Analyses*](#_Samples_Used_for). The sample also excludes students who are assigned either the LOSS or the HOSS. Refer to subsection [*7.2.2 Scale Scores for the Total Assessment*](#_Scale_Scores_for) for definitions of these terms.

#### Reliability Measures

In a specified population of students, the reliability of test scores, *X*, is defined as the proportion of the test score variance that is attributable to true differences in student abilities. It is sometimes operationalized as the correlation between scores on two administrations of the same testing procedure, , where *X* refers to the first administration and *X′* refers to the second administration.

Reliability coefficients may range from 0 to 1. The higher the reliability coefficient for a set of scores, the more likely students would be to obtain very similar scores if they were retested. In applied settings, the requirement of repeated administrations is impractical, and methodologies estimating reliability from relationships among student performances on items within a single test form are often used.

An IRT-based approach called marginal reliability (Green et al., 1984) can be used to estimate the reliability of the scores. The estimates of reliability coefficients reported here are for IRT-based ability estimates.

This reliability coefficient for theta estimates, , is defined on the basis of a single test administration, as shown in equation 8.14. *Refer to the* [*Alternative Text for Equation 8.14*](#_Alternative_Text_for_16) *for a description of this equation.*

 (8.14)

where,

*J* is the number of students who took the assessment,

 is the measure of variance in ability estimates, and

 is the squared conditional standard error of measurement (CSEM) (that is, error variances) for student *j* with ability estimate .

#### Standard Error of Measurement

The SEM is a measure of how much students’ scores would vary from the scores they would earn on a perfectly reliable assessment. If it were possible to compute the error of measurement for each student’s score in a large group of students, these errors of measurement would have a mean of zero. These SEMs are an indication of how much the errors of measurement affect the students’ scores. The SEM is expressed in the same units as the test score, whether the units are in raw score or scale score metric.

The SEM is the square root of the error variance in the scores (that is, the SD of the distribution of the differences between students’ observed scores and their true scores). The SEM is calculated using equation 8.15. *Refer to the* [*Alternative Text for Equation 8.15*](#_Alternative_Text_for_49) *for a description of this equation.*

 (8.15)

where,

 is the reliability estimated in equation 8.14,

 is the SD of the total test theta score, and

*A* is the slope of the scaling transformation of theta scoresto the reporting scale.

The SEM is useful in determining the confidence interval (CI) that likely captures a student’s true score. A student’s true score can be thought of as the mean of observed scores a student would earn over an infinite number of independent administrations of the assessment. Across those administrations, approximately 95 percent of the time the interval ranging from the student’s observed score minus 1.96 SEMs to the student’s observed score plus 1.96 SEMs would contain that student’s true score (Crocker & Algina, 1986). Therefore, this interval is called a 95 percent CI for the student’s true score. For example, if a student’s observed score on a given assessment equals 500 points, and the SEM equals 5, one can be 95 percent confident that the student’s true score lies between 490 and 510 points (500  10).

Table 8.13 provides the total score reliability for theta as well as the mean, SD, and SEM of both thetas and scale scores for each grade level and the high school grade band, along with the number of students upon which those analyses were performed. Note that in the case of the total test reliability, the reliability is for the total test on the theta score scale; it is calculated using the total test theta score of individual students. The test reliability ranged from 0.91 to 0.92 across all grade levels, indicating high levels of reliability.

Table 8.13 Summary Statistics for Scale Scores and Theta Scores, Reliability, and SEMs

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Number of Students** | **Reliability** | **Scale Score Mean** | **Scale Score SD** | **Scale Score SEM** | **Theta Score Mean** | **Theta Score SD** | **Theta Score SEM** |
| Grade 5 | 421,426 | 0.92 | 201 | 22.11 | 5.82 | 0.02 | 1.06 | 0.31 |
| Grade 8 | 425,735 | 0.91 | 401 | 21.97 | 5.83 | –0.01 | 1.06 | 0.31 |
| High school—Grade 10 | 24,907 | 0.91 | 599 | 20.29 | 5.72 | –0.09 | 0.95 | 0.28 |
| High school—Grade 11 | 331,315 | 0.92 | 602 | 21.74 | 5.62 | 0.07 | 1.04 | 0.29 |
| High school—Grade 12 | 99,413 | 0.92 | 599 | 21.82 | 5.65 | –0.08 | 1.04 | 0.29 |
| High school—All grades | 455,635 | 0.92 | 601 | 21.73 | 5.64 | 0.03 | 1.04 | 0.29 |

#### Student Group Reliabilities and Standard Errors of Measurement

The reliabilities of the total assessment scores were also examined for various student groups within the student population. These student groups include demographic student groups, as well as groups of students who took both CAST and the English Language Proficiency Assessments for California (ELPAC). The characteristics considered are gender, ethnicity, economic status, disability status, migrant status, English language fluency, parent/guardian military status, homeless status, and ethnicity by economic status; refer to table 7.17 for the demographic student groups reported.

##### Reliabilities by Demographic Student Groups

The student groups included in these analyses were defined by gender, economic status, disability status, accommodations for students with disabilities, English language fluency, primary ethnicity, migrant status, parent/guardian military status, homeless status, and a crosstab of primary ethnicity and economic status.

Table 8.G.1 through table 8.G.6 in [appendix 8.G](#_Appendix_8.G:_Reliability) provide reliabilities, theta-based SEMs, and theta score variances for the total assessment scores for each student group for each grade level and for high school overall. EL students stood out as having the lowest reliabilities compared to other student groups, which will be discussed in the next subsection, [8.6.4.2](#_Reliabilities_by_ELPAC).

Note that score reliabilities and their associated SEMs for some student groups are presented in the table as “N/A” under a few scenarios:

* Student group with fewer than 11 students
* Score reliabilities that are not estimable (for example, when there is only one student in the student group)
* Negative reliability estimates due to small variation in scale scores and large CSEMs for extreme score values

##### Reliabilities by ELPAC Performance Levels

A subset of students who took CAST also took the Summative ELPAC, which is the required state assessment for English language proficiency (ELP) that must be given to students whose primary language is a language other than English and who are classified as ELs. The Summative ELPAC results show the overall English performance level attained by students. The performance levels for the Summative ELPAC are reported as the following:

* Level 1: Beginning to Develop
* Level 2: Somewhat Developed
* Level 3: Moderately Developed
* Level 4: Well Developed

Detailed descriptions of these ELPAC performance levels can be found on the Summative ELPAC General PLDs web page on the CDE website.

CAST student group reliabilities, as well as SEM and theta score variances, are calculated for all students taking both CAST and the ELPAC, as well as for each of the four ELPAC performance levels. These results show the degree of consistency between low-performance ELPAC scores and the corresponding performance levels, and overall CAST scores. It is anticipated that low reliability estimates are associated with low performance on the ELPAC, and that reliability estimates increase as students demonstrate improved ELP. These results are provided in table 8.G.7 of [appendix 8.G](#_Appendix_8.G:_Reliability).

#### Conditional Standard Errors of Measurement

Classical test theory assumes that the standard error of a test score is constant throughout the score range. While the assumption is probably reasonable in the mid-score ranges, it is less reasonable at the extremes of the score distribution. IRT expands the concept by providing estimates of the standard error at each score point on the distribution.

##### Methodology

CSEMs are estimated as part of the IRT-based scoring procedure. CSEMs for scale scores are based on IRT and are estimated as a function of measured ability. The CSEMs of theta scores (or of linearly transformed theta scores) are smaller at points of the scale in the test metric where more items are located. A student’s CSEM under the IRT framework is equal to the reciprocal of the square root of the test information function (TIF) based on the items taken by each student. The CSEM for a student with proficiency  is calculated using equation 8.16. *Refer to the* [*Alternative Text for Equation 8.16*](#_Alternative_Text_for_50) *for a description of this equation.*

 (8.16)

where,

 is the test information for student *j* and is calculated using equation 8.17. *Refer to the* [*Alternative Text for Equation 8.17*](#_Alternative_Text_for_24) *for a description of this equation.*

 (8.17)

where,

*I* is the number of items on the test form, and

I sub i of theta-hat sub j is the item information of item *i* for student *j*.

Item information is calculated as presented in equation 8.18. *Refer to the* [*Alternative Text for Equation 8.18*](#_Alternative_Text_for_25) *for a description of this equation.*

 (8.18)

where,

is the probability of a student with proficiency  obtaining score *h* on item *i*, the computation of which is shown in equation 8.7; and

*Mi* is the maximum possible score points for item *i*.

CSEMs for scale scores are computed by transforming CSEMs of theta scores onto the reporting scale. Refer to subsection [*8.4.6.2* *Transformation from Theta Scores to Scale Scores*](#_Transformation_from_Theta_1)for scaling procedures.

CAST scaling is different from other California Assessment of Student Performance and Progress (CAASPP) assessments because it involves a nonlinear transformation from theta scores to scale scores. An extra step is conducted to convert the theta score onto an expected raw score on a reference form. The CSEM of the final scale score is calculated as presented in equation 8.19. *Refer to the* [*Alternative Text for Equation 8.19*](#_Alternative_Text_for_29) *for a description of this equation.*

 (8.19)

where,

 is the CSEM on the scale score metric for student *j,*

 is the CSEM on the theta score metric for student *j* estimated in equation 8.16, and

*A\** is the scaling factor (the slope) needed to transform the expected raw score on a reference form to the scale score.

##### Results

As described in subsection [*4.4.2 Test Forms*](#_Test_Forms), CAST has multiple unique test forms, resulting in multiple tables of scale scores, thetas, and CSEMs. For the purpose of this technical report, average CSEMs are presented in figures and tables, for each grade level—grade five, eight, ten, eleven, or twelve—and the high school grade band. [Appendix 8.H](#_Appendix_8.H:_Conditional) provides average CSEM curves and the data used to create these figures. Table 8.14 is a summary of the highlighted information in table 8.H.1 through table 8.H.6, that is, average CSEMs at the thresholds for students to reach achievement levels 2 through 4—*Standard Nearly Met*, *Standard Met*, and *Standard Exceeded*, respectively.

Table 8.14 Average CSEMs at Thresholds for Achievement Levels

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Achievement Levels** | **Scale Score Threshold** | **Average Theta** | **Average Theta CSEM** |
| Grade 5 | Standard Nearly Met | 179 | –0.97 | 0.28 |
| Grade 5 | Standard Met | 214 | 0.51 | 0.27 |
| Grade 5 | Standard Exceeded | 231 | 1.35 | 0.33 |
| Grade 8 | Standard Nearly Met | 378 | –1.01 | 0.30 |
| Grade 8 | Standard Met | 415 | 0.57 | 0.25 |
| Grade 8 | Standard Exceeded | 433 | 1.48 | 0.33 |
| High school—Grade 10 | Standard Nearly Met | 576 | –1.12 | 0.30 |
| High school—Grade 10 | Standard Met | 615 | 0.56 | 0.25 |
| High school—Grade 10 | Standard Exceeded | 636 | 1.67 | 0.32 |
| High school—Grade 11 | Standard Nearly Met | 576 | –1.12 | 0.30 |
| High school—Grade 11 | Standard Met | 615 | 0.56 | 0.25 |
| High school—Grade 11 | Standard Exceeded | 636 | 1.67 | 0.32 |
| High school—Grade 12 | Standard Nearly Met | 576 | –1.12 | 0.30 |
| High school—Grade 12 | Standard Met | 615 | 0.56 | 0.25 |
| High school—Grade 12 | Standard Exceeded | 636 | 1.67 | 0.32 |
| High school—All grades | Standard Nearly Met | 576 | –1.12 | 0.30 |
| High school—All grades | Standard Met | 615 | 0.56 | 0.25 |
| High school—All grades | Standard Exceeded | 636 | 1.67 | 0.32 |

#### Decision Classification Analyses

When an assessment uses achievement levels as the primary method to report test results, accuracy and consistency of decisions become key indicators of the quality of the assessment.

##### Methodology

The reliabilities of achievement-level classifications, which are criterion referenced, are related to the reliabilities of the test scores on which they are based; however, they are not exactly the same. Glaser (1963) was among the first to draw attention to this distinction, and Feldt and Brennan (1989) reviewed the topic extensively. While test reliability evaluates the consistency of test scores, decision classification reliability evaluates the consistency of classification.

Decision accuracy is the extent to which students are classified in the same way as they would be if each student’s score were the average over all possible forms of the assessment (the student’s true score). Decision accuracy answers the following question: How closely does the actual classification of students, 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 could somehow be known?

Decision consistency is the extent to which students are classified in the same way as they would be on the basis of a single form of an assessment other than the one for which data is available. Decision consistency answers the following question: What is the agreement between the classifications based on two nonoverlapping, equally difficult forms of the assessment?

The methodology used for estimating the reliability of classification decisions is described in Livingston and Lewis (1995). The necessary input information includes only the maximum and minimum possible scores on the assessment and the observed score distribution and the reliability coefficient for the group of students referenced by the estimates. The method was implemented by the ETS proprietary computer program RELCLASS-COMP (Version 4.14).

Reliability of classification at a threshold is estimated by combining the achievement levels above a particular threshold and combining the achievement levels below that threshold. The result is a two-by-two table indicating whether the students are above or below the threshold. The sum of the entries in the main diagonal is the number of students accurately (or consistently) classified as above or below that threshold.

Table 8.15 and table 8.16 illustrate these two-by-two contingency tables. The proportion of students being accurately classified is determined by summing across the diagonals of table 8.15. The proportion of consistently classified students is determined by summing the diagonals of table 8.16.

Table . Decision Accuracy for Reaching an Achievement Level

|  |  |  |
| --- | --- | --- |
| Achievement Level Status | Does Not Reach an Achievement Level Based on True Score | Reaches an Achievement Level Based on True Score |
| Does not reach an achievement level | Accurate classification | Inaccurate classification |
| Reaches an achievement level | Inaccurate classification | Accurate classification |

Table 8.16 Decision Consistency for Reaching an Achievement Level

|  |  |  |
| --- | --- | --- |
| **Achievement Level Status** | **Does Not Reach an Achievement Level Based on an Alternate Form** | **Reaches an Achievement Level Based on an Alternate Form** |
| Does not reach an achievement level | Consistent classification | Inconsistent classification |
| Reaches an achievement level | Inconsistent classification | Consistent classification |

##### Results

For an assessment with three threshold scores, the classification is a partition of the distributions of true scores and observed scores into a four-by-four table with the diagonal elements representing accurate or consistent classifications based on the two score distributions. The results of decision accuracy and consistency analysis for CAST are presented, by grade levels and the high school grade band, in table 8.I.1 through table 8.I.12 in [appendix 8.I](#_Appendix_8.I:_Analyses). The proportion of students accurately classified is the sum of the main diagonal elements of the decision accuracy tables. Likewise, the proportion of students consistently classified is the sum of the main diagonal elements of the decision consistency tables.

Using the threshold of Standard Met, the classifications are collapsed to *Standard Not Met* and *Standard Nearly Met* versus *Standard Met* and *Standard Exceeded*, which are the critical categories for accountability. The resulting table is a two-by-two table with diagonal elements representing accurate or consistent classifications. The proportion of students accurately classified in the two collapsed categories is the sum of the main diagonal elements of the collapsed two-by-two decision accuracy tables. Likewise, the proportion of students consistently classified in the two collapsed categories is the sum of the main diagonal elements of the collapsed two-by-two decision consistency tables.

The percentages of students who were classified accurately ranged from 0.80 to 0.83 across all achievement levels and from 0.93 to 0.94 for the two collapsed categories. For decision consistency, the percentages of students classified consistently ranged from 0.73 to 0.76 across all achievement levels and 0.90 to 0.92 for the two collapsed categories.

#### Interrater Agreement

The interrater reliability analyses are performed on approximately 10 percent of the overall testing population, randomly selected from the total population; those students’ responses are scored by two raters. The two sets of ratings are used to compute statistics describing the consistency (or reliability) of the ratings. This interrater consistency is described in four ways:

1. Percentage agreement between two raters
2. Cohen’s Kappa
3. Quadratic-weighted kappa (QWK) coefficient
4. SMD

In some scoring rubrics, zero is a valid score for the responses but is not provided by a rater. Instead, a score of zero is assigned when the student attempted the writing task but did not provide a response. Responses with zero scores should not be included in the calculation of the agreement statistics for these items.

##### Agreement Statistics

###### Percentage Agreement

Percentage agreement between two raters is frequently defined as the percentage of exact score agreement and adjacent score agreement. Exact score agreement means two raters give exact same scores. Adjacent score agreement means agreement between scores that differ by just one point. The percentage of exact score agreement is a stringent criterion, which tends to decrease with an increasing number of possible item score points. The fewer the item score points, the fewer degrees of freedom on which two raters can vary, and the higher the percentage of agreement.

###### Kappa

Interrater reliability or consistency is an indicator of homogeneity and is most frequently measured using Cohen’s Kappa statistic (1960), which takes chance agreement into account. For a human-scored item with *m+1* categories (where *m* is the number of score categories of an item), one can construct an (*m+1)* × *(m+1)* rating table with scores provided by two raters, *X* and *Y*, as shown in table 8.17. Let *nst* denote the number of responses for which rater *X’s* score = *s* and rater *Y’s* score = *t,* *ns+* is the number of responses for which rater *X’s* score = *s*, *n+t* is the number of responses for which rater *Y’s* score = *t*, and *n++* is the number of all responses. An ellipsis (…) signifies that there might be more rows or columns in the table.

Table 8.17 Frequencies of Ratings

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Rating** | **Y = 0** | **Y = 1** | **Y = 2** | **…** | **Y = m** |
| X = 0 | n00 | n01 | n02 | … | n0m |
| X = 1 | N10 | n11 | n12 | … | n1m |
| X = 2 | n20 | n21 | n22 | … | n2m |
| … | … | … | … | … | … |
| X = m | nm0 | nm1 | nm2 | … | nmm |

*Refer to the* [*Alternative Text for Equation 8.20*](#_Alternative_Text_for_40) *for a description of this equation.* The kappa statistic is defined as

 (8.20)

*Refer to the* [*Alternative Text for Equation 8.21*](#_Alternative_Text_for_41) *for a description of this equation.*

 (8.21)

*Refer to the* [*Alternative Text for Equation 8.22*](#_Alternative_Text_for_31) *for a description of this equation.*

 (8.22)

where,

*pobs* is the observed agreement, and

*pexp* is the expected agreement between *X* and *Y*.

When *pobs* and *pexp* agree only at the chance level, the value of kappa is 0. When the two measurements agree perfectly, the value of kappa is 1.0.

###### Quadratic-Weighted Kappa

QWK is used because kappa does not take into account the degree of disagreement between raters. It is a generalization of the simple kappa coefficient using weights to quantify the relative difference between categories. The range of the QWK is from 0.0 to 1.0, with perfect agreement being equal to 1.0.

For a human-scored item with *m+1* categories, one can construct an (*m+1)* × *(m+1)* rating table with scores provided by two raters, *X* and *Y,* as described in table 8.17. The weighted kappa coefficient is defined as presented in equation 8.23. *Refer to the* [*Alternative Text for Equation 8.23*](#_Alternative_Text_for_32) *for a description of this equation.*

 (8.23)

For QWK, the weights are calculated using equation 8.24. *Refer to the* [*Alternative Text for Equation 8.24*](#_Alternative_Text_for_33) *for a description of this equation.*

 (8.24)

###### Standardized Mean Difference

The SMD is an effect size measure used to compare the mean human ratings to the mean AI scores. It is standardized using the SD of the human ratings, since the human raters are the target or the control group. The version of the SMD that uses a target group’s SD and not a pooled SD is called Glass’ Delta (1976). The range of the SMD is from negative to positive infinity, with 0 being the center. An SMD of 0 would indicate no difference in means. A positive value indicates that the AI scoring engine gives higher scores, on average. A negative value indicates that the AI scoring engine gives lower scores, on average. Glass’ Delta (Δ) to compare humans and machine scores is given by equation 8.25. *Refer to the* [*Alternative Text for Equation 8.25*](#_Alternative_Text_for_38) *for a description of this equation.*

 (8.25)

where,

 and  are the mean of the machine scores and mean of the human ratings, respectively, and

 is the SD of the human ratings.

Since the differences are standardized, we may compare SMDs across prompts, even with different score scale lengths. Established criteria to interpret SMDs suggest an absolute value of SMD less than 0.20 to indicate a small or negligible effect (Cohen, 1988; Sawilowsky, 2009). In the AI scoring model evaluation literature, a value of 0.15 has been suggested as a threshold (Williamson et al., 2012) such that the absolute value of SMD (that is, |SMD|) should not exceed 0.15.

###### Proportional Reduction in Mean Squared Error

The proportional reduction in mean squared error (PRMSE) provides an estimate of the relationship between the AI score and the human true score. The PRMSE for predicting a true score 𝑇 with a machine score 𝑀 is defined using equation 8.26. *Refer to the* [*Alternative Text for Equation 8.26*](#_Alternative__Text) *for a description of this equation*.

 (8.26)

The mean squared error (MSE) of a machine score *M* as a predictor of the variable *T* is estimated using equation 8.27. *Refer to the* [*Alternative Text for Equation 8.27*](#_Alternative__) *for a description of this equation*.

 (8.27)

where,

 is the average of the two human scores for examinee *j*,

*Mi* is the machine score for examinee *j*,

*J* is the number of students who have two human scores, and

*VE* is defined using equation 8.28. *Refer to the* [*Alternative Text for Equation 8.28*](#_Alternative_Text_for_51) *for a description of this equation*.

 (8.28)

where,

*Hj1* and *Hj2* are the two observed human scores for examinee *j*, and

*J* is the number of students who have two human scores.

The variance of the true score is approximated with equation 8.29. *Refer to the* [*Alternative Text for Equation 8.29*](#_Alternative_Text_for_54) *for a description of this equation*.

 (8.29)

where,

 is the average of the two human scores for examinee *j*, and

 is the average of all human scores across all examinees.

The PRMSE corrects for the reliability of the human ratings and so provides a clearer measure of AI score prediction accuracy. It can reveal evidence in support of AI scoring even when human ratings have lower reliability. The PRMSE statistic provides an estimate of the proportion of variance in human scores that can be accounted for by automated scores. In the context of kindergarten through grade twelve assessments, values of PRMSE of 0.70 or greater constitute satisfactory evidence for using the AI scores (ETS, 2021).

##### Results of the Interrater Reliabilities

To ensure that the AI scoring engine awarded scores that were consistent with the scores assigned by qualified human raters, ongoing quality checks were conducted to ensure that the scoring models perform consistently. A description of these quality checks is provided in subsection [*9.5.3 Artificial Intelligence Scoring Verification*](#_Artificial_Intelligence_Scoring).

The interrater reliabilities for operational CR items are shown in [appendix 8.J](#_Appendix_8.J:_Interrater). Table 8.J.1 through table 8.J.3 show AI-scored CR items. The QWK ranged from 0.68 to 0.94 for all CR items. These values indicate a moderate to high level of agreement between two raters. Detailed information on interrater reliability results can be found in subsection [*7.1.1.2.9 Interrater Reliability for Operational Items*](#_Interrater_Reliability_for).

### Validity Evidence

Validity refers to the degree to which each interpretation or use of a test score is supported by the accumulated evidence (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014; ETS, 2014). Concerns about validity drive the development, administration, and scoring of an assessment. Validity evidence also determines the appropriateness of test score interpretations and uses.

Validation is the process of accumulating evidence to support each proposed score interpretation or use. This validation process does not rely on a single study or gathering only one type of evidence. Rather, validation involves multiple investigations and different kinds of supporting evidence (AERA, APA, & NCME, 2014; Cronbach, 1971; ETS, 2014; Kane, 2006). It begins with the test design and is implicit throughout the entire assessment process, which includes item development and field testing, analyses of items, standard setting, test scaling and linking, scoring, reporting, and score usage.

In this section, the evidence gathered is presented to support the intended uses and interpretations of scores for CAST. This section discusses some of the principles prescribed by the AERA, APA, and NCME *Standards for Educational and Psychological Testing* (2014). These *Standards* require a clear definition of the purpose of the assessment, a description of the constructs to be assessed, and the population to be assessed, as well as how the scores are to be interpreted and used.

The *Standards* identify five kinds of evidence that can provide support for score interpretations and uses:

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
5. Evidence based on the consequences of testing

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

#### Design of CAST

##### Purpose

CAST is designed to measure performance on the California Next Generation Science Standards (CA NGSS). The goal of CAST is to measure what students know and can do in science. CAST covers information across the three science domains of Earth and Space Sciences, Life Sciences, and Physical Sciences.

##### Constructs to Be Measured

CAST is designed to show how well students perform relative to the CA NGSS. These standards describe what students should know and be able to do at each grade level.

The test blueprint defines the procedures used to measure the domains and standards. The blueprint is provided in table 4.A.1 through table 4.A.6 in [appendix 4.A](#_Appendix_4.A:_Test_1). It also provides an operational definition of the construct to which each set of standards refers. That is, the blueprint defines, for each content area, the subject to be assessed, the tasks to be presented, the administration instructions to be given, and the rules used to score student responses. The test blueprint controls as many aspects of the measurement procedure as possible so that the testing conditions will remain the same over test administrations (Cronbach, 1971) to minimize construct-irrelevant score variance (Messick, 1989).

##### Interpretations and Uses of the Scores

Overall student performance is expressed as scale scores and achievement levels. An inference is drawn about how much knowledge and skill, as measured by CAST, the student has, on the basis of a student’s total score. The total score is also used to classify students in terms of their level of knowledge and skill, on the basis of their performance on the CAST. These levels are called achievement levels and are labeled *Standard Exceeded*, *Standard Met*, *Standard Nearly Met*, and *Standard Not Met*. The descriptions reflecting the level of expectation on students’ performance of these achievement levels can be found in subsection [*7.3.1 Total-Test Achievement Levels*](#_Total-Test_Achievement_Levels). A detailed description of the uses and applications of CAST scores is presented in [*Chapter 7: Scoring and Reporting*](#_Scoring_and_Reporting). Additional information can be found in the *CAASPP and ELPAC Scoring and Reporting Guide* (CDE, 2024).

CAST results have four primary purposes:

1. Help facilitate conversations between parents/guardians and teachers about student performance
2. Serve as a tool to help parents/guardians and teachers work together to improve student learning
3. Help staff from schools and local educational agencies (LEAs) identify strengths and areas that need improvement in their educational programs
4. Provide the public and policymakers with information about student achievement

More detailed descriptions regarding score use can be found in the *Education Code* Section 60602 web page on the California Legislative Information website.

##### Intended Test Population

Students enrolled in grades five and eight are required to take part in CAST, unless they are eligible to participate in the alternate assessments or their parents/guardians have opted them out of testing. Students enrolled in high school are required to take the CAST once while in high school (that is, grade ten, eleven, or twelve [if the student is not repeating grade twelve]), unless they are eligible to participate in the alternate assessments, their parents/guardians have opted them out of testing, or they were not tested because of a medical emergency.

#### Content

Evidence based on test content refers to traditional forms of content validity evidence, such as the rating of test specifications and test items (Crocker, Miller, & Franks, 1989; Sireci, 1998), as well as alignment methods for educational assessments that evaluate the interactions between curriculum frameworks, testing, and instruction (Rothman et al., 2002; Bhola et al., 2003; Martone & Sireci, 2009).

##### Description of California Next Generation Science Standards

As noted in section [*1.1 Background*](#_Background), CAST is aligned with the CA NGSS. There are three main domains at each grade level: Earth and Space Sciences, Life Sciences, and Physical Sciences. Performance expectations (PEs) within the CA NGSS are assessable statements of what students should know and be able to do in each science domain. Overall, the alignment study results provide strong support that the CAST system produces aligned test forms (CDE, 2019).

##### Item Specifications

Item specifications describe the characteristics of items that are written to measure each content standard. Specifications were developed for each PE at each grade level. Details on item specifications can be found in subsection [*3.2.3 Item Specific**ations*](#_Item_Specifications).

##### Assessment Blueprint

The CAST blueprint provided in table 4.A.1 through table 4.A.6 of [appendix 4.A](#_Appendix_4.A:_Test_1) describes the content of the science assessments for all grade levels tested and how that content is assessed. The CAST blueprint reflects the depth and breadth of the PEs of the CA NGSS. The test blueprint has information about the number of items and depth of knowledge for items associated with each assessment target. Each assessment is described by a single blueprint for each segment of the assessment. For details about the CAST blueprint, refer to subsection [*4.3.1 Test Blu**eprint*](#_Test_Blueprint).

##### Alignment Study

A strong alignment between CAST and the CA NGSS is fundamental to the meaningful measurement of student achievement and instructional effectiveness. Alignment results demonstrate that CAST represents the full range of the content standards and measures student knowledge in the same manner and at the same level of complexity as expected in the content standards. For detailed information on the alignment study conducted, refer to the *California Science Test (CAST) Alignment Study Report* (CDE, 2019).

##### Form Assembly Process

The content standards and blueprint are the basis for choosing items for each assessment. Assembly of all CAST forms meets all the content and statistical specifications. Refer to section[*4.4 Test Production Process*](#_Test_Production_Process) for additional information.

#### Response Processes

Validity evidence based on response processes refers to “evidence concerning the fit between the construct and the detailed nature of performance or response actually engaged in by test takers” (AERA et al., 2014, p. 15). This type of evidence can include documentation of activities such as

* systematic observations of test response behavior; and
* student responses to survey questions concerning the test items on the test form;
* evaluation of the criteria used by judges when scoring CRs, analysis of student item response time data, and features scored by automated algorithms (Embretson, 1983; Messick, 1989; Mislevy, 2009).

This type of evidence is used to confirm that the CAST is measuring the cognitive skills that are intended to be the objects of measurement and that students are using these targeted skills to respond to the items.

##### Testing Time Analysis

Testing times for each administration can be evaluated for consistency, with the expected response processes for the tasks presented to students. The length of time it takes students to take an assessment is recorded and analyzed to build a profile describing what a typical testing event looks like for each segment, item type, content domain and grade level. In addition, variability in testing time is investigated to determine whether a student’s testing time should be viewed as unusual or irregular. It should be noted that CAST assessments are untimed.

Response time analyses were based on students who logged on to the assessment and whose total testing time at the test level did not equal zero. The descriptive statistics—for example, the number of students, mean, SD, minimum and maximum, and percentiles—of the segment, item type, and content domain are computed for each grade level. Refer to section [*8.5 Response Time Analyses*](#_Response_Time_Analyses) and [appendix 8.F](#_Appendix_8.F:_Response) for detailed testing time analyses and results.

##### Student Survey

The student survey questions were administered at the end of the assessment and focused on gathering information about how the science content on the CAST compared to the science content presented in the classroom. There were four survey questions. The survey questions asked how many topics on the assessment were taught in the students’ science classes, whether the test questions are different from the types of questions in their science class, and how hard the questions on this assessment compared to questions in their science class. For grades ten and eleven students in high school only, the survey asked whether the student would be enrolling in any more science classes.

The student survey results show that, in general, CAST reflects what students were taught in the classroom. Detailed information on the student survey can be found in [*Cha**p**te**r 10: Student Survey*](#_Student_Survey_1).

Survey questions were provided in braille for students who used the braille accommodation.

#### Internal Structure

Evidence based on *internal structure* refers to the statistical analysis of item and score subdomains to investigate the primary and secondary (if any) dimensions measured by an assessment. A dimensionality study was conducted for CAST on the basis of 2018–19 test data.

Analysis of the internal structure evidence also includes indices of measurement precision such as DIF analyses, test reliability, student group reliability, decision accuracy and consistency, interrater agreement, conditional and unconditional SEMs, and TIFs.

##### Dimensionality

CAST assesses PEs as they appear in the CA NGSS, and the PEs represent a complete integration of the three dimensions (that is, SEP, DCI, and CCC), not three dimensions that coincide together. A dimensionality study was conducted during the 2018–19 test administration to determine the factorial structure of the assessments. Results suggested the assessment is essentially unidimensional, which is consistent with the notion of the CAST design in that it measures the integration of the dimensions. Details on the dimensionality study can be found in chapter 12 of the *California Science Test 2018–2019 Technical Report* (CDE, 2020).

##### Differential Item Functioning

Analysis of item functioning using DIF falls under the internal structure category. For CAST, DIF analyses are conducted to assess differences in the item performance of groups of students who differ in their demographic characteristics. For the 2023–24 CAST administration, two items in grade five, one item in grade eight, and one item in high school were identified as having significant levels of DIF, and none of the items were considered biased. The details on how DIF analyses are performed can be found in section [*8.3 Differential Item Functioning Analyses*](#_Differential_Item_Functioning).

##### Overall Reliability Estimates

The results of marginal reliability analyses on the total theta scores for CAST are presented in table 8.13. Also provided in table 8.13 are the mean, SD, and SEM of both thetas and scale scores for each grade level and the high school grade band, along with the number of students whose data was used to perform those analyses. Note that in the case of the total test reliability, the reliability is for the total assessment on the theta score scale; it is calculated using the total test theta score of individual students. The results indicate that the reliability estimates for the CAST total scores are high, ranging from 0.91 to 0.92 across all grade levels.

#### Relations to Other Variables

Evidence based on *relations to other variables* refers to traditional forms of criterion-related validity evidence such as concurrent and predictive validity, as well as more comprehensive investigations of the relationships among test scores and other variables such as multitrait-multimethod studies (Campbell & Fiske, 1959). External variables can be used to evaluate hypothesized relationships between test scores and other measures of student achievement (for example, test scores on other assessments) to evaluate the degree to which different assessments actually measure different skills and the utility of test scores for predicting specific criteria (for example, college grades). This type of evidence is essential for supporting the validity of certain inferences based on CAST scores.

Most students from grades five, eight, and eleven who take the CAST also take the Smarter Balanced English language arts/literacy (ELA) and mathematics assessments. Table 8.K.1 through table 8.K.6 in [appendix 8.K](#_Appendix_8.K:_Correlations) show these correlations for both ELA and mathematics test scores for all students and by demographic student groups. Table 8.K.1 and table 8.K.2 show data for grade five, table 8.K.3 and table 8.K.4 show data for grade eight, and table 8.K.5 and table 8.K.6 show data for grade eleven—there is no data for grades ten and twelve, because these grade levels did not participate in Smarter Balanced testing. For the total student group, across all grade levels, similar correlations were observed for both content areas. The correlations between the CAST and Smarter Balanced for ELA were 0.84 for grade five, 0.81 for grade eight, and 0.78 for grade eleven. The correlations between the CAST and Smarter Balanced for Mathematics were 0.81 for grade five, 0.82 for grade eight, and 0.78 for grade eleven.

Table 8.L.1 through table 8.L.6 in [appendix 8.L](#_Appendix_8.L:_Correlations) present correlations between CAST domain scores and total test scores; and Smarter Balanced composite claim scores and total test scores for ELA and mathematics, for grades five, eight, and eleven. For all three grade levels, moderate, positive correlations were observed between the Smarter Balanced composite claim scores and the CAST domain scores. The correlations between the Smarter Balanced for ELA Reading and Listening claims and the CAST domains were about the same as the correlations between the Writing and Research claims and the CAST domains. The correlations between the Smarter Balanced for Mathematics Concepts and Procedures claim scores and the CAST domain scores were about the same as the Mathematical Practices (Problem Solving, Communicating Reasoning, and Modeling and Data Analysis) claim scores, for all three grade levels.

Table 8.18 presents the total number of students who took the CAST, as well as the matched and unmatched percentages. *Matched Percentage* represents the percentage of students who also took the ELPAC. *Not Matched Percentage* includes the percentage of students who took the CAST but did not take the ELPAC. Grade five, grade eight, and grades ten through twelve are shown in this table because these are the grade levels at which the ELPAC is administered. The relatively low matched percentage indicates that only a subset of CAST test takers took the ELPAC, because most CAST test takers are not English learners. The demographic student group composition of the samples is shown in [appendix 7.C](#_Appendix_7.C:_Demographic), table 7.C.1 through table 7.C.6.

Table 8.18 also presents the correlations between CAST and ELPAC scores; correlations between 0.43 and 0.62 are observed. A science assessment typically requires the use of less English language than an English language assessment; therefore, it is reasonable that correlations between the CAST and ELPAC scores are of a moderate, positive magnitude.

Table 8.18 Correlations Between CAST and ELPAC Scores

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grade Level** | **Number of Students Taking CAST** | **Matched Percentage** | **Not Matched Percentage** | **Number of Students Taking Both** | **Correlation** |
| Grade 5 | 422,398 | 19 | 81 | 80,348 | 0.62 |
| Grade 8 | 427,448 | 13 | 87 | 54,340 | 0.45 |
| Grade 10 | 24,935 | 12 | 88 | 2,992 | 0.43 |
| Grade 11 | 331,949 | 10 | 90 | 34,057 | 0.45 |
| Grade 12 | 99,602 | 12 | 88 | 11,468 | 0.44 |

#### Consequences of Testing

Evidence based on *consequences of testing* refers to the evaluation of the intended and unintended consequences associated with a testing program. Examples of evidence based on testing consequences include investigations of adverse impact, evaluation of the effects of testing on instruction, and evaluation of the effects of testing studies on issues such as high school dropout rates. With respect to educational assessments, the *Standards* stress the importance of evaluating test consequences. For example, they state the following:

When educational testing programs are mandated…the ways in which test results are intended to be used should be clearly described by those who mandate the tests. It is also the responsibility of those who mandate the use of tests to monitor their impact and to identify and minimize potential negative consequences as feasible. Consequences resulting from the uses of the test, both intended and unintended, should also be examined by the test developer and/or user. (AERA et al., 2014, p. 195)

Investigations of testing consequences relevant to the CAST goals include analyses of students’ opportunity to learn the CA NGSS and analyses of changes in textbooks and instructional approaches. Unintended consequences, such as diminished morale among teachers and students, increased pressure on students leading to increased dropout rates, or the pursuit of college majors and careers that are less challenging can be evaluated. These sorts of investigations require information beyond what has been available to the CAST program to date.

### Special Considerations for the Spanish-Translation Version

In past test administrations, when the number of students taking the Spanish-translation version of the CAST exceeded 500, the item analyses would be conducted for items in the Spanish-translation version, and the procedure is the same as the analyses for the general forms. However, starting from the 2022–23 test administration, new methods were implemented to effectively identify items in the Spanish version to ensure that the Spanish translation and the rendering of the stacked Spanish items in the TDS did not cause the items to perform significantly differently from their general version counterparts.

Two methods were used to evaluate how the performance of items in the Spanish-translation version compared to their counterparts in the general version. The first method provides the *p*-values and the item-total correlations for the items in the Spanish-translation version and identifies items that significantly deviate from the general trend of the comparison of items in the two forms (that is, outside two standard deviations [SDs] from the average difference). The second method is a DIF approach that compares the item performance across students taking the two forms who have been matched on their overall test performance. These two methods provide different information on the comparison and were both useful in understanding how the items in the Spanish-translation version compared to the general version.

Items identified with both approaches were sent to the content experts for further review of the Spanish translation, item rendering, and potential bias. ETS Assessment Development (AD) staff reviewed 13 grade five items, 11 grade eight items, and 12 high school items in 2023–‍24, and no issue was found. Table 8.19 presents the number of items by domain across the two grade levels and the high school grade band. Note that the flagged items need to be reviewed before ETS Psychometric Analysis & Research (PAR) group starts the IRT calibration.

Table 8.19 Spanish Item Analyses Domain Distribution

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** |
| Grade 5 | 4 | 3 | 6 |
| Grade 8 | 5 | 0 | 6 |
| High school | 5 | 5 | 2 |

Table 8.20 presents the distribution of the item types across the two grade levels and the high school grade band. Note that a composite item is defined as an item type that includes multiple parts, MC refers to a multiple-choice item, and TEI refers to a technology-enhanced item.

Table 8.20 Spanish Item Analyses Item Type Distribution

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Composite** | **MC** | **TEI** |
| Grade 5 | 4 | 1 | 8 |
| Grade 8 | 1 | 5 | 5 |
| High school | 1 | 6 | 5 |

### References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states’ content standards: Methods and issues. *Educational Measurement: Issues and Practice*, *22*, 21–‍29.

Cai, L. (2017). FlexMIRT®: *Flexible multilevel, multidimensional item analysis and test scoring* (Version 3.5) [computer software]. Chapel Hill, NC: Vector Psychometric Group.

California Department of Education. (2019)*.* *California Science Test (CAST) alignment study report* [Unpublished report]. Sacramento, CA: California Department of Education.

California Department of Education. (2020). *California Science Test: 2018–2019 technical report*. California Department of Education website.

California Department of Education. (2024). *CAASPP and ELPAC scoring and reporting guide*. Sacramento, CA: California Department of Education.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

Cohen, J. (1988). *Statistical power analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Crocker, L. M., Miller, D., & Franks, E. A. (1989). Quantitative methods for assessing the fit between test and curriculum. *Applied Measurement in Education*, *2*, 179–94.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

Dorans, N. J. (2013). ETS contributions to the quantitative assessment of item, test, and score fairness. *ETS Research Report Series*, i–38.

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., & 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.

Educational Testing Service. (2014). *ETS standards for quality and fairness*. Princeton, NJ: Educational Testing Service.

Embretson (Whitley), S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, *93*, 179–97.

ETS. (2021). *Best practices for constructed-response scoring*. Princeton, NJ: ETS.

Feldt, L. S. & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd edition) (pp. 105–46). New York: Macmillan.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist*,18, 519–32.

Glass, G. V (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 38.

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, *21*(4), 347–60.

Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65–110). Washington, DC: American Council on Education and National Council on Measurement in Education.

Holland, P. W., & Thayer, D. T. (1985). *An alternative definition of the ETS delta scale of item difficulty* (Research Report 85–43). Princeton, NJ: Educational Testing Service.

Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Erlbaum.

Huynh, H. (2000, June). *Guidelines for Rasch linking for PACT*. Memorandum to Paul Sandifer on June 18, 2000. Columbia, SC: Available from Author.

Huynh, H., & Rawls, A. (2009). A comparison between robust z and 0.3-logit difference procedures in assessing stability of linking items for the Rasch model. In Everett V. Smith Jr. & Greg E. Stone (Eds.) *Applications of Rasch Measurement in Criterion-Reference Testing: Practice Analysis to Score Reporting.* Maple Grove, MN: JAM Press.

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

Kim, S., & Kolen M. J. (2004). STUIRT: A Computer Program for Scale Transformation under Unidimensional Item Response Theory Models. University of Iowa. Version 1.0.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, linking, and scaling: Methods and practices* (2nd ed.). New York, NY: Springer-Verlag.

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.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–‍74.

Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessments, and instruction.*Review of Educational Research*, *4*, 1332–61.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed.). Washington, DC: American Council on Education.

Mislevy, R. J. (2009). Validity from the perspective of model-based reasoning. *CRESST Report 752*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

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

Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika*, *47*, 337–47.

Rothman, R., Slattery, J. B., Vranek, J. L., & Resnick, L. B. (2002). *Benchmarking and alignment of standards and testing* [Technical Report 566]. Washington, DC: Center for the Study of Evaluation.

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 26.

Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment*, *5*, 299–321.

Stocking, M. L. (1996). An alternative method for scoring adaptive tests. *Journal of Educational and Behavioral Statistics*, *21*, 365–89.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*(2), 201–10.

Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice*, *31*(1), 2–13.

Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). Descriptive and inferential procedures for assessing differential item functioning in polytomous items. *Applied Measurement in Education*, *10*(4), 321–44.

### Accessibility Information

#### Alternative Text for Equation 8.1

p value sub dich equals the fraction with the numerator the sum from j equals 1 to J sub i of X sub ij and the denominator J sub i end fraction. *(Return to* [*equation 8.1*](#EQ8_1)*.)*

#### Alternative Text for Equation 8.2

p value sub poly equals the fraction with the numerator the sum from j equals 1 to J sub i of X sub ij and the denominator J sub i times M sub i end fraction. *(Return to* [*equation 8.2*](#EQ8_2)*.)*

#### Alternative Text for Equation 8.3

r sub polyreg equals the fraction beta-hat times s sub tot divided by the square root of beta-hat squared times s squared sub tot plus 1. *(Return to* [*equation 8.3*](#EQ8_3)*.)*

#### Alternative Text for Equation 8.4

alpha sub MH equals the numerator open parenthesis the sum from m equals 1 to M of R sub rm times W sub fm divided by N sub tm close parenthesis divided by the denominator open parenthesis the sum from m equals 1 to M of R sub fm times W sub rm divided by N sub tm closed parenthesis. *(Return to* [*equation 8.4*](#EQ8_4)*.)*

#### Alternative Text for Equation 8.5

MH D-DIF equals negative 2.35 times the natural logarithm open bracket alpha sub MH close bracket. *(Return to* [*equation 8.5*](#EQ8_5)*.)*

#### Alternative Text for Equation 8.6

SMD equals the fraction with numerator the sum from m equals 1 to M of N sub fm times E sub fm and denominator the sum from m equals 1 to M of N sub fm end fraction minus the fraction with numerator the sum from m equals 1 to M of N sub fm times E sub rm and denominator the sum from m equals 1 to M of N sub fm end fraction equals the fraction with the numerator the sum from m equals 1 to M of D sub fm and the denominator m equals1 to M of N sub fm end fraction. *(Return to* [*equation 8.6*](#EQ8_6)*.)*

#### Alternative Text for Equation 8.7

p sub ih of theta-hat sub j equals the numerator exp open parenthesis the sum from v equals 1 to h of D times a sub i open parenthesis theta-hat sub j minus b sub i plus d sub iv close parenthesis close parenthesis divided by the denominator open parenthesis 1 plus the sum from c equals 1 to m sub i exp open parenthesis the sum from v equals 1 to c of D times a sub i open parenthesis theta-hat sub j minus b sub i plus d sub iv close parenthesis close parenthesis close parenthesis, if score h equals 1, 2, …, n sub i.

p sub ih of theta-hat sub j equals 1 divided by the denominator open parenthesis 1 plus the sum from c equals 1 to m sub i exp open parenthesis the sum from v equals 1 to c of D times a sub i open parenthesis theta-hat sub j minus b sub i plus d sub iv close parenthesis close parenthesis close parenthesis, if score h equals 0. *(Return to* [*equation 8.7*](#EQ8_7)*.)*

#### Alternative Text for Equation 8.8

d squared sub i equals the sum across k quadrature points the squared difference in the expected item score for item i based on the new parameters and bank parameters multiplied by weights. *(Return to* [*equ**ation 8.8*](#EQ8_8)*.)*

#### Alternative Text for Equation 8.9

epsilon of theta equals the sum from i equals 1 to I of sum from h equals 0 to M sub i of h times p sub ih of theta-hat sub j. *(Return to* [*equation 8.9*](#EQ8_9)*.)*

#### Alternative Text for Equation 8.10

Scale score sub j equals slope multiplied by adjusted number-right sub j plus intercept. *(Return to* [*equation 8.10*](#EQ8_10)*.)*

#### Alternative Text for Equation 8.11

Lowest obtainable scale score plus one equals slope multiplied by lowest obtainable adjusted number-right plus intercept. Highest obtainable scale score equals slope multiplied by lowest obtainable adjusted number-right plus intercept. *(Return to* [*equation 8.11*](#EQ8_11)*.)*

#### Alternative Text for Equation 8.12

Lowest obtainable number-right equals one hundred multiplied by the fraction with the numerator exponent lowest obtainable theta and the denominator one plus lowest obtainable theta. *(Return to* [*equation 8.12*](#EQ8_12)*.)*

#### Alternative Text for Equation 8.13

Highest obtainable number-right equals one hundred multiplied by the fraction with the numerator exponent highest obtainable theta and the denominator one plus highest obtainable theta. *(Return to* [*equation 8.13*](#EQ8_13)*.)*

#### Alternative Text for Equation 8.14

rho sub theta-hat prime equals 1 minus the fraction with the numerator sum from j equals 1 to J of CSEM squared sub theta-hat sub j divided by the denominator J times s squared sub theta-hat. *(Return to* [*equation 8.14*](#EQ8_14)*.)*

#### Alternative Text for Equation 8.15

SEM sub scaled equals A times s sub theta-hat times the square root of 1 minus rho sub theta-hat prime. *(Return to* [*equation 8.15*](#EQ8_15)*.)*

#### Alternative Text for Equation 8.16

CSEM of theta-hat sub j equals 1 divided by the square root of I of theta sub j. *(Return to* [*equation 8.16*](#EQ8_16)*.)*

#### Alternative Text for Equation 8.17

I of theta-hat sub j equals the sum from i equals 1 to I of I sub i of theta-hat sub j. *(Return to* [*equation 8.17*](#EQ8_17)*.)*

#### Alternative Text for Equation 8.18

I sub i of theta-hat sub j equals the sum from h equals 0 to M sub I of h squared times p sub ih of theta-hat sub j minus open bracket sum from h equals 0 to M sub i of h times p sub ih of theta-hat sub j close bracket squared. *(Return to* [*equation 8.18*](#EQ8_18)*.)*

#### Alternative Text for Equation 8.19

CSEM of SS sub j equals slope times CSEM of theta-hat sub j times 100 times the numerator exp theta-hat sub j divided by the denominator open parenthesis 1 plus exp theta-hat sub j close parenthesis times 1 divided by the denominator open parenthesis 1 plus exp theta-hat sub j close parenthesis. *(Return to* [*equation 8.19*](#EQ8_19)*.)*

#### Alternative Text for Equation 8.20

kappa equals the fraction with the numerator p sub obs minus p sub exp the denominator 1 minus p sub exp. *(Return to* [*equation 8.20*](#EQ8_20)*.)*

#### Alternative Text for Equation 8.21

P sub obs equals 1 divided by n times the sum from s equals 0 to m n sub ss. *(Return to* [*equation 8.21*](#EQ8_21)*.)*

#### Alternative Text for Equation 8.22

P sub exp equals 1 divided by n square times the sum from s equals 0 to m n sub s plus times n sub plus s. *(Return to* [*equation 8.22*](#EQ8_22)*.)*

#### Alternative Text for Equation 8.23

K sub ij equals open parenthesis the sum from i equals zero to m the sum from j equals zero to m of w sub ij times n sub ij divided by n sub plus plus close parenthesis minus open parenthesis the sum from i equals zero to m the sum from j equals zero to m of w sub ij times n sub I plus times n sub plus j divided by n squared sub plus plus close parenthesis divided open parenthesis 1 minus open parenthesis the sum from i equals zero to m the sum from j equals zero to m of w sub ij times n sub i plus times n sub plus j divided by n squared sub plus plus close parenthesis close parenthesis, K sub ij equals open parenthesis the sum from i equals zero to m the sum from j equals zero to m of w sub ij times n sub ij divided by n sub plus plus close parenthesis minus open parenthesis the sum from i equals zero to m the sum from j equals zero to m of w sub ij times n sub i plus times n sub plus j divided by n squared sub plus plus close parenthesis divided open parenthesis 1 minus open parenthesis the sum from i equals zero to m the sum from j equals zero to m of w sub ij times n sub i plus times n sub plus j divided by n squared sub plus plus close parenthesis close parenthesis. *(Return to* [*equation 8.23*](#EQ8_23)*.)*

#### Alternative Text for Equation 8.24

W sub ij equals 1 minus open parenthesis I minus j close parenthesis squared divided by m squared. *(Return to* [*equation 8.24*](#EQ8_24)*.)*

#### Alternative Text for Equation 8.25

Delta equals open parenthesis the mean of machine scores minus the mean of human ratings close parenthesis divided by the standard deviation of human ratings. *(Return to* [*equation 8.25*](#EQ8_25)*.)*

#### Alternative Text for Equation 8.26

PRMSE of T given E equals to 1 minus MSE of T given M divided by variance of T. *(Return to* [*equation 8.26*](#EQ8_26)*.)*

#### Alternative Text for Equation 8.27

MSE of T given M equals to the sum from j equals 1 to J of open bracket H bar sub j minus M sub j close bracket squared divided by J minus one half times V sub E. *(Return to* [*equation 8.27*](#EQ8_27)*.)*

#### Alternative Text for Equation 8.28

V sub E equals 1 divided by 2 times J the sum from j equals 1 to J of open bracket H sub j1 minus H sub j2 close bracket squared.*(Return to* [*equation 8.28*](#EQ8_28)*.)*

#### Alternative Text for Equation 8.29

Variance of T equals to the sum from j equals 1 to J of open bracket H bar sub j minus H bar close bracket squared divided by open bracket J minus one close bracket minus one half times V sub E.*(Return to* [*equation 8.29*](#EQ8_29)*.)*

### Appendix 8.A: Test-Taking Rates

**Notes:**

* This set of tables shows the percentage of registered students in selected demographic student groups who tested.
* The total numbers of registered students are derived from version 3 of the production data file (“P3”).
* A student is considered a test taker if the student was enrolled during the active testing window and logged on to the test.
* High school grades are ten, eleven, and twelve.

Table 8.A. CAST Test-Taking Rates for Grade Five by Student Group

|  |  |  |  |
| --- | --- | --- | --- |
| Student Group | Number of Registered Students | Number Tested | Percent Tested |
| All students | 429,092 | 422,600 | 98.5 |
| Male | 219,152 | 215,649 | 98.4 |
| Female | 209,818 | 206,835 | 98.6 |
| Nonbinary | 122 | 116 | 95.1 |
| EL | 82,017 | 80,711 | 98.4 |
| English only | 264,396 | 259,751 | 98.2 |
| Reclassified fluent English proficient (RFEP) | 55,788 | 55,516 | 99.5 |
| Initial fluent English proficient (IFEP) | 26,665 | 26,489 | 99.3 |
| Adult English learner (ADEL) | 0 | 0 | 0.0 |
| To be determined | 140 | 70 | 50.0 |
| English proficiency unknown | 86 | 63 | 73.3 |
| Economically disadvantaged | 281,278 | 277,454 | 98.6 |
| Not economically disadvantaged | 147,814 | 145,146 | 98.2 |
| American Indian or Alaska Native | 1,874 | 1,808 | 96.5 |
| Asian | 45,336 | 44,962 | 99.2 |
| Native Hawaiian or Other Pacific Islander | 1,742 | 1,710 | 98.2 |
| Filipino | 9,477 | 9,413 | 99.3 |
| Hispanic or Latino | 238,039 | 235,068 | 98.8 |
| Black or African American | 21,438 | 20,911 | 97.5 |
| White | 86,357 | 84,451 | 97.8 |
| Two or more races | 24,829 | 24,277 | 97.8 |
| Disability | 58,883 | 56,624 | 96.2 |
| No disability | 370,209 | 365,976 | 98.9 |
| Migrant education | 3,481 | 3,443 | 98.9 |
| Not migrant education | 425,611 | 419,157 | 98.5 |
| Armed forces family member | 6,592 | 6,494 | 98.5 |
| Not armed forces family member | 422,500 | 416,106 | 98.5 |
| Homeless | 16,004 | 15,591 | 97.4 |
| Not homeless | 413,088 | 407,009 | 98.5 |
| Foster youth | 1,412 | 1,355 | 96.0 |
| Not foster youth | 427,680 | 421,245 | 98.5 |

Table 8.A. CAST Test-Taking Rates for Grade Eight by Student Group

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Group** | **Number of Registered Students** | **Number Tested** | **Percent Tested** |
| All students | 441,145 | 428,080 | 97.0 |
| Male | 225,714 | 219,345 | 97.2 |
| Female | 214,998 | 208,351 | 96.9 |
| Nonbinary | 433 | 384 | 88.7 |
| EL | 56,268 | 54,510 | 96.9 |
| English only | 253,081 | 243,558 | 96.2 |
| RFEP | 113,078 | 111,684 | 98.8 |
| IFEP | 18,534 | 18,234 | 98.4 |
| ADEL | 0 | 0 | 0.0 |
| To be determined | 121 | 58 | 47.9 |
| English proficiency unknown | 63 | 36 | 57.1 |
| Economically disadvantaged | 286,629 | 278,715 | 97.2 |
| Not economically disadvantaged | 154,516 | 149,365 | 96.7 |
| American Indian or Alaska Native | 1,946 | 1,812 | 93.1 |
| Asian | 42,998 | 42,498 | 98.8 |
| Native Hawaiian or Other Pacific Islander | 1,925 | 1,857 | 96.5 |
| Filipino | 10,172 | 10,079 | 99.1 |
| Hispanic or Latino | 249,286 | 242,869 | 97.4 |
| Black or African American | 22,480 | 21,511 | 95.7 |
| White | 89,506 | 85,576 | 95.6 |
| Two or more races | 22,832 | 21,878 | 95.8 |
| Disability | 54,562 | 50,819 | 93.1 |
| No disability | 386,583 | 377,261 | 97.6 |
| Migrant education | 3,394 | 3,352 | 98.8 |
| Not migrant education | 437,751 | 424,728 | 97.0 |
| Armed forces family member | 6,161 | 5,997 | 97.3 |
| Not armed forces family member | 434,984 | 422,083 | 97.0 |
| Homeless | 14,450 | 13,816 | 95.6 |
| Not homeless | 426,695 | 414,264 | 97.1 |
| Foster youth | 1,611 | 1,439 | 89.3 |
| Not foster youth | 439,534 | 426,641 | 97.1 |

Table 8.A. CAST Test-Taking Rates for Grade Ten by Student Group

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Group** | **Number of Registered Students** | **Number Tested** | **Percent Tested** |
| All students | 29,080 | 24,975 | 85.9 |
| Male | 14,998 | 12,836 | 85.6 |
| Female | 14,005 | 12,074 | 86.2 |
| Nonbinary | 77 | 65 | 84.4 |
| EL | 3,686 | 3,079 | 83.5 |
| English only | 15,973 | 13,778 | 86.3 |
| RFEP | 8,251 | 7,085 | 85.9 |
| IFEP | 1,112 | 986 | 88.7 |
| ADEL | 45 | 41 | 91.1 |
| To be determined | 1 | 0 | 0.0 |
| English proficiency unknown | 12 | 6 | 50.0 |
| Economically disadvantaged | 19,592 | 16,747 | 85.5 |
| Not economically disadvantaged | 9,488 | 8,228 | 86.7 |
| American Indian or Alaska Native | 139 | 130 | 93.5 |
| Asian | 1,827 | 1,639 | 89.7 |
| Native Hawaiian or Other Pacific Islander | 98 | 84 | 85.7 |
| Filipino | 605 | 560 | 92.6 |
| Hispanic or Latino | 18,397 | 15,406 | 83.7 |
| Black or African American | 1,579 | 1,466 | 92.8 |
| White | 5,245 | 4,681 | 89.2 |
| Two or more races | 1,190 | 1,009 | 84.8 |
| Disability | 3,804 | 3,102 | 81.5 |
| No disability | 25,276 | 21,873 | 86.5 |
| Migrant education | 255 | 233 | 91.4 |
| Not migrant education | 28,825 | 24,742 | 85.8 |
| Armed forces family member | 404 | 314 | 77.7 |
| Not armed forces family member | 28,676 | 24,661 | 86.0 |
| Homeless | 815 | 761 | 93.4 |
| Not homeless | 28,265 | 24,214 | 85.7 |
| Foster youth | 174 | 151 | 86.8 |
| Not foster youth | 28,906 | 24,824 | 85.9 |

Table 8.A. CAST Test-Taking Rates for Grade Eleven by Student Group

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Group** | **Number of Registered Students** | **Number Tested** | **Percent Tested** |
| All students | 355,094 | 332,507 | 93.6 |
| Male | 181,678 | 170,195 | 93.7 |
| Female | 172,649 | 161,635 | 93.6 |
| Nonbinary | 767 | 677 | 88.3 |
| EL | 38,827 | 35,224 | 90.7 |
| English only | 192,749 | 179,147 | 92.9 |
| RFEP | 106,992 | 102,382 | 95.7 |
| IFEP | 16,430 | 15,703 | 95.6 |
| ADEL | 0 | 0 | 0.0 |
| To be determined | 62 | 26 | 41.9 |
| English proficiency unknown | 34 | 25 | 73.5 |
| Economically disadvantaged | 221,257 | 207,073 | 93.6 |
| Not economically disadvantaged | 133,837 | 125,434 | 93.7 |
| American Indian or Alaska Native | 1,598 | 1,428 | 89.4 |
| Asian | 34,668 | 33,332 | 96.1 |
| Native Hawaiian or Other Pacific Islander | 1,534 | 1,414 | 92.2 |
| Filipino | 8,935 | 8,634 | 96.6 |
| Hispanic or Latino | 200,011 | 188,065 | 94.0 |
| Black or African American | 16,268 | 14,742 | 90.6 |
| White | 75,501 | 69,680 | 92.3 |
| Two or more races | 16,579 | 15,212 | 91.8 |
| Disability | 40,665 | 35,738 | 87.9 |
| No disability | 314,429 | 296,769 | 94.4 |
| Migrant education | 2,378 | 2,256 | 94.9 |
| Not migrant education | 352,716 | 330,251 | 93.6 |
| Armed forces family member | 5,160 | 4,974 | 96.4 |
| Not armed forces family member | 349,934 | 327,533 | 93.6 |
| Homeless | 11,327 | 10,160 | 89.7 |
| Not homeless | 343,767 | 322,347 | 93.8 |
| Foster youth | 1,329 | 1,139 | 85.7 |
| Not foster youth | 353,765 | 331,368 | 93.7 |

Table 8.A. CAST Test-Taking Rates for Grade Twelve by Student Group

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Group** | **Number of Registered Students** | **Number Tested** | **Percent Tested** |
| All students | 121,646 | 99,829 | 82.1 |
| Male | 62,236 | 51,051 | 82.0 |
| Female | 59,103 | 48,566 | 82.2 |
| Nonbinary | 307 | 212 | 69.1 |
| EL | 15,656 | 12,075 | 77.1 |
| English only | 63,770 | 50,992 | 80.0 |
| RFEP | 35,275 | 31,131 | 88.3 |
| IFEP | 6,252 | 5,382 | 86.1 |
| ADEL | 390 | 207 | 53.1 |
| To be determined | 77 | 10 | 13.0 |
| English proficiency unknown | 226 | 32 | 14.2 |
| Economically disadvantaged | 75,062 | 61,898 | 82.5 |
| Not economically disadvantaged | 46,584 | 37,931 | 81.4 |
| American Indian or Alaska Native | 639 | 459 | 71.8 |
| Asian | 12,512 | 11,150 | 89.1 |
| Native Hawaiian or Other Pacific Islander | 663 | 532 | 80.2 |
| Filipino | 3,237 | 2,884 | 89.1 |
| Hispanic or Latino | 67,780 | 56,208 | 82.9 |
| Black or African American | 7,333 | 5,557 | 75.8 |
| White | 23,611 | 18,537 | 78.5 |
| Two or more races | 5,871 | 4,502 | 76.7 |
| Disability | 15,573 | 10,974 | 70.5 |
| No disability | 106,073 | 88,855 | 83.8 |
| Migrant education | 929 | 820 | 88.3 |
| Not migrant education | 120,717 | 99,009 | 82.0 |
| Armed forces family member | 1,583 | 1,406 | 88.8 |
| Not armed forces family member | 120,063 | 98,423 | 82.0 |
| Homeless | 5,352 | 3,985 | 74.5 |
| Not homeless | 116,294 | 95,844 | 82.4 |
| Foster youth | 818 | 508 | 62.1 |
| Not foster youth | 120,828 | 99,321 | 82.2 |

Table 8.A. CAST Test-Taking Rates for High School (All Grades Tested) by Student Group

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Group** | **Number of Registered Students** | **Number Tested** | **Percent Tested** |
| All students | 505,820 | 457,311 | 90.4 |
| Male | 258,912 | 234,082 | 90.4 |
| Female | 245,757 | 222,275 | 90.4 |
| Nonbinary | 1,151 | 954 | 82.9 |
| EL | 58,169 | 50,378 | 86.6 |
| English only | 272,492 | 243,917 | 89.5 |
| RFEP | 150,518 | 140,598 | 93.4 |
| IFEP | 23,794 | 22,071 | 92.8 |
| ADEL | 435 | 248 | 57.0 |
| To be determined | 140 | 36 | 25.7 |
| English proficiency unknown | 272 | 63 | 23.2 |
| Economically disadvantaged | 315,911 | 285,718 | 90.4 |
| Not economically disadvantaged | 189,909 | 171,593 | 90.4 |
| American Indian or Alaska Native | 2,376 | 2,017 | 84.9 |
| Asian | 49,007 | 46,121 | 94.1 |
| Native Hawaiian or Other Pacific Islander | 2,295 | 2,030 | 88.5 |
| Filipino | 12,777 | 12,078 | 94.5 |
| Hispanic or Latino | 286,188 | 259,679 | 90.7 |
| Black or African American | 25,180 | 21,765 | 86.4 |
| White | 104,357 | 92,898 | 89.0 |
| Two or more races | 23,640 | 20,723 | 87.7 |
| Disability | 60,042 | 49,814 | 83.0 |
| No disability | 445,778 | 407,497 | 91.4 |
| Migrant education | 3,562 | 3,309 | 92.9 |
| Not migrant education | 502,258 | 454,002 | 90.4 |
| Armed forces family member | 7,147 | 6,694 | 93.7 |
| Not armed forces family member | 498,673 | 450,617 | 90.4 |
| Homeless | 17,494 | 14,906 | 85.2 |
| Not homeless | 488,326 | 442,405 | 90.6 |
| Foster youth | 2,321 | 1,798 | 77.5 |
| Not foster youth | 503,499 | 455,513 | 90.5 |

**Note:** The grade twelve cohort in table 8.A.7 includes students who tested in grades ten, eleven, or twelve and are enrolled in grade twelve in 2023–24.

Table 8.A. CAST Test-Taking Rates for the Grade Twelve Cohort by Student Group

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Group** | **Number of Registered Students** | **Number Tested** | **Percent Tested** |
| All students | 460,182 | 433,135 | 94.1 |
| Male | 234,279 | 220,819 | 94.3 |
| Female | 224,947 | 211,484 | 94.0 |
| Nonbinary | 956 | 832 | 87.0 |
| EL | 45,262 | 41,349 | 91.4 |
| English only | 246,025 | 229,282 | 93.2 |
| RFEP | 145,428 | 140,516 | 96.6 |
| IFEP | 22,676 | 21,645 | 95.5 |
| ADEL | 484 | 299 | 61.8 |
| To be determined | 80 | 11 | 13.8 |
| English proficiency unknown | 227 | 33 | 14.5 |
| Economically disadvantaged | 268,986 | 253,629 | 94.3 |
| Not economically disadvantaged | 191,196 | 179,506 | 93.9 |
| American Indian or Alaska Native | 2,063 | 1,845 | 89.4 |
| Asian | 45,044 | 43,368 | 96.3 |
| Native Hawaiian or Other Pacific Islander | 1,981 | 1,834 | 92.6 |
| Filipino | 12,543 | 12,125 | 96.7 |
| Hispanic or Latino | 259,317 | 245,789 | 94.8 |
| Black or African American | 22,429 | 20,367 | 90.8 |
| White | 96,399 | 89,133 | 92.5 |
| Two or more races | 20,406 | 18,674 | 91.5 |
| Disability | 49,497 | 43,672 | 88.2 |
| No disability | 410,685 | 389,463 | 94.8 |
| Migrant education | 3,029 | 2,907 | 96.0 |
| Not migrant education | 457,153 | 430,228 | 94.1 |
| Armed forces family member | 6,886 | 6,659 | 96.7 |
| Not armed forces family member | 453,296 | 426,476 | 94.1 |
| Homeless | 15,283 | 13,778 | 90.2 |
| Not homeless | 444,899 | 419,357 | 94.3 |
| Foster youth | 1,934 | 1,606 | 83.0 |
| Not foster youth | 458,248 | 431,529 | 94.2 |

### Appendix 8.B: Item Difficulty Distribution

**Note:**

Item types are as follows:

* MC = multiple-choice item.
* CR = constructed-response item.
* TEI = technology-enhanced item.
* Composite = composite item (an item type that includes multiple parts).

Table 8.B.1 Item Difficulty Distributions by Item Type

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Item Type** | **0 ≤ p < 0.2** | **0.2 ≤ p < 0.4** | **0.4 ≤ p < 0.6** | **0.6 ≤ p < 0.8** | **0.8 ≤ p ≤ 1.0** | **Total Number of Items** |
| Grade 5 | MC | 1 | 4 | 17 | 10 | 0 | 32 |
| Grade 5 | CR | 1 | 3 | 4 | 0 | 0 | 8 |
| Grade 5 | TEI | 0 | 4 | 7 | 6 | 1 | 18 |
| Grade 5 | Composite | 0 | 0 | 2 | 1 | 0 | 3 |
| Grade 8 | MC | 0 | 9 | 16 | 1 | 0 | 26 |
| Grade 8 | CR | 0 | 4 | 4 | 1 | 0 | 9 |
| Grade 8 | TEI | 0 | 12 | 9 | 4 | 0 | 25 |
| Grade 8 | Composite | 0 | 0 | 1 | 0 | 0 | 1 |
| High school—Grade 10 | MC | 1 | 9 | 18 | 2 | 0 | 30 |
| High school—Grade 10 | CR | 2 | 6 | 1 | 0 | 0 | 9 |
| High school—Grade 10 | TEI | 2 | 9 | 10 | 5 | 0 | 26 |
| High school—Grade 10 | Composite | 0 | 1 | 1 | 0 | 0 | 2 |
| High school—Grade 11 | MC | 1 | 6 | 20 | 3 | 0 | 30 |
| High school—Grade 11 | CR | 1 | 6 | 2 | 0 | 0 | 9 |
| High school—Grade 11 | TEI | 2 | 6 | 12 | 5 | 1 | 26 |
| High school—Grade 11 | Composite | 0 | 0 | 2 | 0 | 0 | 2 |
| High school—Grade 12 | MC | 1 | 9 | 17 | 3 | 0 | 30 |
| High school—Grade 12 | CR | 1 | 7 | 1 | 0 | 0 | 9 |
| High school—Grade 12 | TEI | 2 | 8 | 11 | 5 | 0 | 26 |
| High school—Grade 12 | Composite | 0 | 1 | 1 | 0 | 0 | 2 |
| High school—All grades | MC | 1 | 7 | 19 | 3 | 0 | 30 |
| High school—All grades | CR | 1 | 7 | 1 | 0 | 0 | 9 |
| High school—All grades | TEI | 2 | 7 | 11 | 6 | 0 | 26 |
| High school—All grades | Composite | 0 | 1 | 1 | 0 | 0 | 2 |

**Note:**

Content domains are as follows:

* ESS = Earth and Space Sciences.
* LS = Life Sciences.
* PS = Physical Sciences.

Table 8.B.2 Item Difficulty Distributions by Content Domain

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Content Domain** | **0 ≤ p < 0.2** | **0.2 ≤ p < 0.4** | **0.4 ≤ p < 0.6** | **0.6 ≤ p < 0.8** | **0.8 ≤ p ≤ 1.0** | **Total Number of Items** |
| Grade 5 | ESS | 0 | 3 | 11 | 6 | 0 | 20 |
| Grade 5 | LS | 1 | 3 | 11 | 6 | 0 | 21 |
| Grade 5 | PS | 1 | 5 | 8 | 5 | 1 | 20 |
| Grade 8 | ESS | 0 | 7 | 9 | 2 | 0 | 18 |
| Grade 8 | LS | 0 | 9 | 11 | 2 | 0 | 22 |
| Grade 8 | PS | 0 | 9 | 10 | 2 | 0 | 21 |
| High school—Grade 10 | ESS | 1 | 6 | 12 | 2 | 0 | 21 |
| High school—Grade 10 | LS | 3 | 9 | 9 | 2 | 0 | 23 |
| High school—Grade 10 | PS | 1 | 10 | 9 | 3 | 0 | 23 |
| High school—Grade 11 | ESS | 1 | 5 | 13 | 2 | 0 | 21 |
| High school—Grade 11 | LS | 3 | 4 | 13 | 2 | 1 | 23 |
| High school—Grade 11 | PS | 0 | 9 | 10 | 4 | 0 | 23 |
| High school—Grade 12 | ESS | 1 | 6 | 12 | 2 | 0 | 21 |
| High school—Grade 12 | LS | 3 | 8 | 10 | 2 | 0 | 23 |
| High school—Grade 12 | PS | 0 | 11 | 8 | 4 | 0 | 23 |
| High school—All grades | ESS | 1 | 5 | 13 | 2 | 0 | 21 |
| High school—All grades | LS | 3 | 7 | 10 | 3 | 0 | 23 |
| High school—All grades | PS | 0 | 10 | 9 | 4 | 0 | 23 |

### Appendix 8.C: Item-Total Correlation Distribution

**Note:**

Item types are as follows:

* MC = multiple-choice item.
* CR = constructed-response item.
* TEI = technology-enhanced item.
* Composite = composite item (an item type that includes multiple parts).

Table 8.C.1 Item-Total Correlation Distributions by Item Type

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Item Type** | **r < 0** | **0 ≤ r < 0.2** | **0.2 ≤ r < 0.3** | **0.3 ≤ r < 0.4** | **0.4 ≤ r < 0.5** | **r ≥ 0.5** | **Total Number of Items** |
| Grade 5 | MC | 0 | 0 | 0 | 1 | 8 | 23 | 32 |
| Grade 5 | CR | 0 | 0 | 0 | 0 | 0 | 8 | 8 |
| Grade 5 | TEI | 0 | 0 | 0 | 2 | 3 | 13 | 18 |
| Grade 5 | Composite | 0 | 0 | 0 | 0 | 2 | 1 | 3 |
| Grade 8 | MC | 0 | 1 | 2 | 4 | 4 | 15 | 26 |
| Grade 8 | CR | 0 | 0 | 0 | 0 | 0 | 9 | 9 |
| Grade 8 | TEI | 0 | 0 | 1 | 3 | 4 | 17 | 25 |
| Grade 8 | Composite | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| High school—Grade 10 | MC | 0 | 0 | 2 | 8 | 10 | 10 | 30 |
| High school—Grade 10 | CR | 0 | 0 | 0 | 0 | 0 | 9 | 9 |
| High school—Grade 10 | TEI | 0 | 0 | 5 | 2 | 4 | 15 | 26 |
| High school—Grade 10 | Composite | 0 | 0 | 0 | 0 | 1 | 1 | 2 |
| High school—Grade 11 | MC | 0 | 0 | 2 | 4 | 11 | 13 | 30 |
| High school—Grade 11 | CR | 0 | 0 | 0 | 0 | 0 | 9 | 9 |
| High school—Grade 11 | TEI | 0 | 0 | 2 | 4 | 5 | 15 | 26 |
| High school—Grade 11 | Composite | 0 | 0 | 0 | 0 | 0 | 2 | 2 |
| High school—Grade 12 | MC | 0 | 0 | 2 | 3 | 13 | 12 | 30 |
| High school—Grade 12 | CR | 0 | 0 | 0 | 0 | 0 | 9 | 9 |
| High school—Grade 12 | TEI | 0 | 0 | 3 | 4 | 5 | 14 | 26 |
| High school—Grade 12 | Composite | 0 | 0 | 0 | 0 | 0 | 2 | 2 |
| High school—All grades | MC | 0 | 0 | 2 | 4 | 11 | 13 | 30 |
| High school—All grades | CR | 0 | 0 | 0 | 0 | 0 | 9 | 9 |
| High school—All grades | TEI | 0 | 0 | 2 | 4 | 5 | 15 | 26 |
| High school—All grades | Composite | 0 | 0 | 0 | 0 | 0 | 2 | 2 |

**Note:**

Content domains are as follows:

* ESS = Earth and Space Sciences.
* LS = Life Sciences.
* PS = Physical Sciences.

Table 8.C.2 Item-Total Correlation Distributions by Content Domain

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Content Domain** | **r < 0** | **0 ≤ r < 0.2** | **0.2 ≤ r < 0.3** | **0.3 ≤ r < 0.4** | **0.4 ≤ r < 0.5** | **r ≥ 0.5** | **Total Number of Items** |
| Grade 5 | ESS | 0 | 0 | 0 | 0 | 3 | 17 | 20 |
| Grade 5 | LS | 0 | 0 | 0 | 2 | 4 | 15 | 21 |
| Grade 5 | PS | 0 | 0 | 0 | 1 | 6 | 13 | 20 |
| Grade 8 | ESS | 0 | 0 | 2 | 3 | 3 | 10 | 18 |
| Grade 8 | LS | 0 | 0 | 0 | 1 | 2 | 19 | 22 |
| Grade 8 | PS | 0 | 1 | 1 | 3 | 3 | 13 | 21 |
| High school—Grade 10 | ESS | 0 | 0 | 3 | 3 | 5 | 10 | 21 |
| High school—Grade 10 | LS | 0 | 0 | 2 | 2 | 6 | 13 | 23 |
| High school—Grade 10 | PS | 0 | 0 | 2 | 5 | 4 | 12 | 23 |
| High school—Grade 11 | ESS | 0 | 0 | 2 | 2 | 5 | 12 | 21 |
| High school—Grade 11 | LS | 0 | 0 | 2 | 1 | 6 | 14 | 23 |
| High school—Grade 11 | PS | 0 | 0 | 0 | 5 | 5 | 13 | 23 |
| High school—Grade 12 | ESS | 0 | 0 | 3 | 1 | 5 | 12 | 21 |
| High school—Grade 12 | LS | 0 | 0 | 2 | 1 | 7 | 13 | 23 |
| High school—Grade 12 | PS | 0 | 0 | 0 | 5 | 6 | 12 | 23 |
| High school—All grades | ESS | 0 | 0 | 2 | 2 | 5 | 12 | 21 |
| High school—All grades | LS | 0 | 0 | 2 | 1 | 6 | 14 | 23 |
| High school—All grades | PS | 0 | 0 | 0 | 5 | 5 | 13 | 23 |

### Appendix 8.D: Item Discrimination Parameter Distribution

**Note:**

Item types are as follows:

* MC = multiple-choice item.
* CR = constructed-response item.
* TEI = technology-enhanced item.
* Composite = composite item (an item type that includes multiple parts).

Table 8.D.1 Item Discrimination Parameter Distribution by Item Type for Grade Five

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **IRT-*a* Range** | **MC** | **CR** | **TEI** | **Composite** | **Number of Items** |
| a < 0 | 0 | 0 | 0 | 0 | 0 |
| 0 ≤ a < 0.2 | 0 | 0 | 0 | 0 | 0 |
| 0.2 ≤ a < 0.4 | 1 | 0 | 3 | 2 | 6 |
| 0.4 ≤ a < 0.6 | 10 | 2 | 6 | 1 | 19 |
| 0.6 ≤ a < 0.8 | 10 | 0 | 5 | 0 | 15 |
| 0.8 ≤ a < 1.0 | 9 | 4 | 3 | 0 | 16 |
| 1.0 ≤ a < 1.2 | 1 | 1 | 1 | 0 | 3 |
| 1.2 ≤ a < 1.4 | 1 | 1 | 0 | 0 | 2 |
| 1.4 ≤ a < 1.6 | 0 | 0 | 0 | 0 | 0 |
| 1.6 ≤ a < 1.8 | 0 | 0 | 0 | 0 | 0 |
| 1.8 ≤ a < 2.0 | 0 | 0 | 0 | 0 | 0 |
| a ≥ 2.0 | 0 | 0 | 0 | 0 | 0 |
| Minimum | 0.30 | 0.46 | 0.30 | 0.31 | 0.30 |
| Maximum | 1.24 | 1.27 | 1.02 | 0.46 | 1.27 |
| Mean | 0.69 | 0.86 | 0.65 | 0.38 | 0.68 |
| SD | 0.21 | 0.26 | 0.21 | 0.08 | 0.23 |
| **Number of Items:** | **32** | **8** | **18** | **3** | **61** |

Table 8.D.2 Item Discrimination Parameter Distribution by Item Type for Grade Eight

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **IRT-*a* Range** | **MC** | **CR** | **TEI** | **Composite** | **Number of Items** |
| a < 0 | 0 | 0 | 0 | 0 | 0 |
| 0 ≤ a < 0.2 | 1 | 0 | 1 | 0 | 2 |
| 0.2 ≤ a < 0.4 | 8 | 0 | 2 | 0 | 10 |
| 0.4 ≤ a < 0.6 | 5 | 0 | 10 | 1 | 16 |
| 0.6 ≤ a < 0.8 | 4 | 1 | 7 | 0 | 12 |
| 0.8 ≤ a < 1.0 | 7 | 2 | 4 | 0 | 13 |
| 1.0 ≤ a < 1.2 | 1 | 6 | 0 | 0 | 7 |
| 1.2 ≤ a < 1.4 | 0 | 0 | 1 | 0 | 1 |
| 1.4 ≤ a < 1.6 | 0 | 0 | 0 | 0 | 0 |
| 1.6 ≤ a < 1.8 | 0 | 0 | 0 | 0 | 0 |
| 1.8 ≤ a < 2.0 | 0 | 0 | 0 | 0 | 0 |
| a ≥ 2.0 | 0 | 0 | 0 | 0 | 0 |
| Minimum | 0.14 | 0.68 | 0.15 | 0.50 | 0.14 |
| Maximum | 1.04 | 1.14 | 1.30 | 0.50 | 1.30 |
| Mean | 0.59 | 0.99 | 0.62 | 0.50 | 0.66 |
| SD | 0.26 | 0.14 | 0.23 | N/A | 0.27 |
| **Number of Items:** | **26** | **9** | **25** | **1** | **61** |

Table 8.D.3 Item Discrimination Parameter Distribution by Item Type for High School

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **IRT-*a* Range** | **MC** | **CR** | **TEI** | **Composite** | **Number of Items** |
| a < 0 | 0 | 0 | 0 | 0 | 0 |
| 0 ≤ a < 0.2 | 0 | 0 | 1 | 0 | 1 |
| 0.2 ≤ a < 0.4 | 6 | 0 | 5 | 0 | 11 |
| 0.4 ≤ a < 0.6 | 15 | 0 | 7 | 1 | 23 |
| 0.6 ≤ a < 0.8 | 3 | 3 | 5 | 1 | 12 |
| 0.8 ≤ a < 1.0 | 4 | 4 | 2 | 0 | 10 |
| 1.0 ≤ a < 1.2 | 2 | 2 | 3 | 0 | 7 |
| 1.2 ≤ a < 1.4 | 0 | 0 | 1 | 0 | 1 |
| 1.4 ≤ a < 1.6 | 0 | 0 | 2 | 0 | 2 |
| 1.6 ≤ a < 1.8 | 0 | 0 | 0 | 0 | 0 |
| 1.8 ≤ a < 2.0 | 0 | 0 | 0 | 0 | 0 |
| a ≥ 2.0 | 0 | 0 | 0 | 0 | 0 |
| Minimum | 0.24 | 0.62 | 0.20 | 0.47 | 0.20 |
| Maximum | 1.12 | 1.17 | 1.55 | 0.65 | 1.55 |
| Mean | 0.58 | 0.91 | 0.69 | 0.56 | 0.67 |
| SD | 0.24 | 0.19 | 0.38 | 0.13 | 0.31 |
| **Number of Items:** | **30** | **9** | **26** | **2** | **67** |

Table 8.D.4 Item Discrimination Parameter Distribution by Content Domain for Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **IRT-*a* Range** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** | **Number of Items** |
| a < 0 | 0 | 0 | 0 | 0 |
| 0 ≤ a < 0.2 | 0 | 0 | 0 | 0 |
| 0.2 ≤ a < 0.4 | 1 | 2 | 3 | 6 |
| 0.4 ≤ a < 0.6 | 7 | 5 | 7 | 19 |
| 0.6 ≤ a < 0.8 | 3 | 7 | 5 | 15 |
| 0.8 ≤ a < 1.0 | 7 | 6 | 3 | 16 |
| 1.0 ≤ a < 1.2 | 1 | 1 | 1 | 3 |
| 1.2 ≤ a < 1.4 | 1 | 0 | 1 | 2 |
| 1.4 ≤ a < 1.6 | 0 | 0 | 0 | 0 |
| 1.6 ≤ a < 1.8 | 0 | 0 | 0 | 0 |
| 1.8 ≤ a < 2.0 | 0 | 0 | 0 | 0 |
| a ≥ 2.0 | 0 | 0 | 0 | 0 |
| Minimum | 0.40 | 0.30 | 0.30 | 0.30 |
| Maximum | 1.24 | 1.02 | 1.27 | 1.27 |
| Mean | 0.73 | 0.67 | 0.65 | 0.68 |
| SD | 0.24 | 0.20 | 0.25 | 0.23 |
| **Number of Items:** | **20** | **21** | **20** | **61** |

Table 8.D.5 Item Discrimination Parameter Distribution by Content Domain for Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **IRT-*a* Range** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** | **Number of Items** |
| a < 0 | 0 | 0 | 0 | 0 |
| 0 ≤ a < 0.2 | 1 | 0 | 1 | 2 |
| 0.2 ≤ a < 0.4 | 5 | 1 | 4 | 10 |
| 0.4 ≤ a < 0.6 | 5 | 6 | 5 | 16 |
| 0.6 ≤ a < 0.8 | 3 | 6 | 3 | 12 |
| 0.8 ≤ a < 1.0 | 2 | 6 | 5 | 13 |
| 1.0 ≤ a < 1.2 | 2 | 2 | 3 | 7 |
| 1.2 ≤ a < 1.4 | 0 | 1 | 0 | 1 |
| 1.4 ≤ a < 1.6 | 0 | 0 | 0 | 0 |
| 1.6 ≤ a < 1.8 | 0 | 0 | 0 | 0 |
| 1.8 ≤ a < 2.0 | 0 | 0 | 0 | 0 |
| a ≥ 2.0 | 0 | 0 | 0 | 0 |
| Minimum | 0.15 | 0.36 | 0.14 | 0.14 |
| Maximum | 1.05 | 1.30 | 1.14 | 1.30 |
| Mean | 0.58 | 0.74 | 0.65 | 0.66 |
| SD | 0.28 | 0.22 | 0.29 | 0.27 |
| **Number of Items:** | **18** | **22** | **21** | **61** |

Table 8.D.6 Item Discrimination Parameter Distribution by Content Domain for High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **IRT-*a* Range** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** | **Number of Items** |
| a < 0 | 0 | 0 | 0 | 0 |
| 0 ≤ a < 0.2 | 1 | 0 | 0 | 1 |
| 0.2 ≤ a < 0.4 | 3 | 3 | 5 | 11 |
| 0.4 ≤ a < 0.6 | 7 | 7 | 9 | 23 |
| 0.6 ≤ a < 0.8 | 4 | 6 | 2 | 12 |
| 0.8 ≤ a < 1.0 | 3 | 2 | 5 | 10 |
| 1.0 ≤ a < 1.2 | 2 | 4 | 1 | 7 |
| 1.2 ≤ a < 1.4 | 0 | 1 | 0 | 1 |
| 1.4 ≤ a < 1.6 | 1 | 0 | 1 | 2 |
| 1.6 ≤ a < 1.8 | 0 | 0 | 0 | 0 |
| 1.8 ≤ a < 2.0 | 0 | 0 | 0 | 0 |
| a ≥ 2.0 | 0 | 0 | 0 | 0 |
| Minimum | 0.20 | 0.24 | 0.32 | 0.20 |
| Maximum | 1.51 | 1.30 | 1.55 | 1.55 |
| Mean | 0.65 | 0.69 | 0.66 | 0.67 |
| SD | 0.33 | 0.30 | 0.32 | 0.31 |
| **Number of Items:** | **21** | **23** | **23** | **67** |

### Appendix 8.E: Item Difficulty Parameter Distribution

**Note:**

Item types are as follows:

* MC = multiple-choice item.
* CR = constructed-response item.
* TEI = technology-enhanced item.
* Composite = composite item (an item type that includes multiple parts).

Table 8.E.1 Item Difficulty Parameter Distribution by Item Type for Grade Five

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **IRT-*b* Range** | **MC** | **CR** | **TEI** | **Composite** | **Number of Items** |
| b < −3.5 | 0 | 0 | 0 | 0 | 0 |
| −3.5 ≤ b < −3.0 | 0 | 0 | 0 | 0 | 0 |
| −3.0 ≤ b < −2.5 | 0 | 0 | 0 | 0 | 0 |
| −2.5 ≤ b < −2.0 | 0 | 0 | 0 | 0 | 0 |
| −2.0 ≤ b < −1.5 | 0 | 0 | 1 | 0 | 1 |
| −1.5 ≤ b < −1.0 | 2 | 0 | 2 | 1 | 5 |
| −1.0 ≤ b < −0.5 | 7 | 0 | 3 | 0 | 10 |
| −0.5 ≤ b < 0 | 7 | 3 | 6 | 1 | 17 |
| 0 ≤ b < 0.5 | 10 | 1 | 2 | 1 | 14 |
| 0.5 ≤ b < 1.0 | 5 | 2 | 2 | 0 | 9 |
| 1.0 ≤ b < 1.5 | 0 | 2 | 1 | 0 | 3 |
| 1.5 ≤ b < 2.0 | 1 | 0 | 1 | 0 | 2 |
| 2.0 ≤ b < 2.5 | 0 | 0 | 0 | 0 | 0 |
| 2.5 ≤ b < 3.0 | 0 | 0 | 0 | 0 | 0 |
| 3.0 ≤ b < 3.5 | 0 | 0 | 0 | 0 | 0 |
| b ≥ 3.5 | 0 | 0 | 0 | 0 | 0 |
| Minimum | −1.15 | −0.42 | −1.63 | −1.18 | −1.63 |
| Maximum | 1.98 | 1.47 | 1.50 | 0.10 | 1.98 |
| Mean | −0.07 | 0.41 | −0.17 | −0.42 | −0.05 |
| SD | 0.65 | 0.73 | 0.85 | 0.67 | 0.73 |
| **Number of Items:** | **32** | **8** | **18** | **3** | **61** |

Table 8.E.2 Item Difficulty Parameter Distribution by Item Type for Grade Eight

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **IRT-*b* Range** | **MC** | **CR** | **TEI** | **Composite** | **Number of Items** |
| b < −3.5 | 0 | 0 | 0 | 0 | 0 |
| −3.5 ≤ b < −3.0 | 0 | 0 | 0 | 0 | 0 |
| −3.0 ≤ b < −2.5 | 0 | 0 | 0 | 0 | 0 |
| −2.5 ≤ b < −2.0 | 0 | 0 | 0 | 0 | 0 |
| −2.0 ≤ b < −1.5 | 0 | 0 | 0 | 0 | 0 |
| −1.5 ≤ b < −1.0 | 0 | 0 | 0 | 0 | 0 |
| −1.0 ≤ b < −0.5 | 1 | 0 | 3 | 0 | 4 |
| −0.5 ≤ b < 0 | 6 | 1 | 7 | 0 | 14 |
| 0 ≤ b < 0.5 | 10 | 5 | 2 | 1 | 18 |
| 0.5 ≤ b < 1.0 | 5 | 2 | 4 | 0 | 11 |
| 1.0 ≤ b < 1.5 | 3 | 1 | 6 | 0 | 10 |
| 1.5 ≤ b < 2.0 | 1 | 0 | 3 | 0 | 4 |
| 2.0 ≤ b < 2.5 | 0 | 0 | 0 | 0 | 0 |
| 2.5 ≤ b < 3.0 | 0 | 0 | 0 | 0 | 0 |
| 3.0 ≤ b < 3.5 | 0 | 0 | 0 | 0 | 0 |
| b ≥ 3.5 | 0 | 0 | 0 | 0 | 0 |
| Minimum | −0.74 | −0.38 | −0.88 | 0.30 | −0.88 |
| Maximum | 1.80 | 1.03 | 1.90 | 0.30 | 1.90 |
| Mean | 0.37 | 0.41 | 0.46 | 0.30 | 0.41 |
| SD | 0.62 | 0.46 | 0.86 | N/A | 0.70 |
| **Number of Items:** | **26** | **9** | **25** | **1** | **61** |

Table 8.E.3 Item Difficulty Parameter Distribution by Item Type for High School

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **IRT-*b* Range** | **MC** | **CR** | **TEI** | **Composite** | **Number of Items** |
| b < −3.5 | 0 | 0 | 0 | 0 | 0 |
| −3.5 ≤ b < −3.0 | 0 | 0 | 0 | 0 | 0 |
| −3.0 ≤ b < −2.5 | 0 | 0 | 0 | 0 | 0 |
| −2.5 ≤ b < −2.0 | 0 | 0 | 0 | 0 | 0 |
| −2.0 ≤ b < −1.5 | 0 | 0 | 1 | 0 | 1 |
| −1.5 ≤ b < −1.0 | 0 | 0 | 2 | 0 | 2 |
| −1.0 ≤ b < −0.5 | 3 | 0 | 2 | 0 | 5 |
| −0.5 ≤ b < 0 | 10 | 0 | 4 | 1 | 15 |
| 0 ≤ b < 0.5 | 6 | 2 | 7 | 0 | 15 |
| 0.5 ≤ b < 1.0 | 7 | 3 | 5 | 1 | 16 |
| 1.0 ≤ b < 1.5 | 0 | 3 | 1 | 0 | 4 |
| 1.5 ≤ b < 2.0 | 3 | 1 | 1 | 0 | 5 |
| 2.0 ≤ b < 2.5 | 1 | 0 | 0 | 0 | 1 |
| 2.5 ≤ b < 3.0 | 0 | 0 | 1 | 0 | 1 |
| 3.0 ≤ b < 3.5 | 0 | 0 | 2 | 0 | 2 |
| b ≥ 3.5 | 0 | 0 | 0 | 0 | 0 |
| Minimum | −0.76 | 0.13 | −1.92 | −0.22 | −1.92 |
| Maximum | 2.26 | 1.94 | 3.16 | 0.56 | 3.16 |
| Mean | 0.35 | 0.92 | 0.45 | 0.17 | 0.46 |
| SD | 0.76 | 0.54 | 1.22 | 0.55 | 0.94 |
| **Number of Items:** | **30** | **9** | **26** | **2** | **67** |

Table 8.E.4 Item Difficulty Parameter Distribution by Content Domain for Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **IRT-*b* Range** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** | **Number of Items** |
| b < −3.5 | 0 | 0 | 0 | 0 |
| −3.5 ≤ b < −3.0 | 0 | 0 | 0 | 0 |
| −3.0 ≤ b < −2.5 | 0 | 0 | 0 | 0 |
| −2.5 ≤ b < −2.0 | 0 | 0 | 0 | 0 |
| −2.0 ≤ b < −1.5 | 0 | 0 | 1 | 1 |
| −1.5 ≤ b < −1.0 | 1 | 2 | 2 | 5 |
| −1.0 ≤ b < −0.5 | 4 | 3 | 3 | 10 |
| −0.5 ≤ b < 0 | 5 | 9 | 3 | 17 |
| 0 ≤ b < 0.5 | 7 | 2 | 5 | 14 |
| 0.5 ≤ b < 1.0 | 1 | 4 | 4 | 9 |
| 1.0 ≤ b < 1.5 | 1 | 1 | 1 | 3 |
| 1.5 ≤ b < 2.0 | 1 | 0 | 1 | 2 |
| 2.0 ≤ b < 2.5 | 0 | 0 | 0 | 0 |
| 2.5 ≤ b < 3.0 | 0 | 0 | 0 | 0 |
| 3.0 ≤ b < 3.5 | 0 | 0 | 0 | 0 |
| b ≥ 3.5 | 0 | 0 | 0 | 0 |
| Minimum | −1.03 | −1.15 | −1.63 | −1.63 |
| Maximum | 1.50 | 1.37 | 1.98 | 1.98 |
| Mean | −0.02 | −0.11 | −0.03 | −0.05 |
| SD | 0.66 | 0.63 | 0.91 | 0.73 |
| **Number of Items:** | **20** | **21** | **20** | **61** |

Table 8.E.5 Item Difficulty Parameter Distribution by Content Domain for Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **IRT-*b* Range** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** | **Number of Items** |
| b < −3.5 | 0 | 0 | 0 | 0 |
| −3.5 ≤ b < −3.0 | 0 | 0 | 0 | 0 |
| −3.0 ≤ b < −2.5 | 0 | 0 | 0 | 0 |
| −2.5 ≤ b < −2.0 | 0 | 0 | 0 | 0 |
| −2.0 ≤ b < −1.5 | 0 | 0 | 0 | 0 |
| −1.5 ≤ b < −1.0 | 0 | 0 | 0 | 0 |
| −1.0 ≤ b < −0.5 | 1 | 2 | 1 | 4 |
| −0.5 ≤ b < 0 | 4 | 4 | 6 | 14 |
| 0 ≤ b < 0.5 | 4 | 8 | 6 | 18 |
| 0.5 ≤ b < 1.0 | 2 | 5 | 4 | 11 |
| 1.0 ≤ b < 1.5 | 6 | 2 | 2 | 10 |
| 1.5 ≤ b < 2.0 | 1 | 1 | 2 | 4 |
| 2.0 ≤ b < 2.5 | 0 | 0 | 0 | 0 |
| 2.5 ≤ b < 3.0 | 0 | 0 | 0 | 0 |
| 3.0 ≤ b < 3.5 | 0 | 0 | 0 | 0 |
| b ≥ 3.5 | 0 | 0 | 0 | 0 |
| Minimum | −0.74 | −0.88 | −0.62 | −0.88 |
| Maximum | 1.90 | 1.74 | 1.80 | 1.90 |
| Mean | 0.53 | 0.31 | 0.42 | 0.41 |
| SD | 0.75 | 0.65 | 0.72 | 0.70 |
| **Number of Items:** | **18** | **22** | **21** | **61** |

Table 8.E.6 Item Difficulty Parameter Distribution by Content Domain for High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **IRT-*b* Range** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** | **Number of Items** |
| b < −3.5 | 0 | 0 | 0 | 0 |
| −3.5 ≤ b < −3.0 | 0 | 0 | 0 | 0 |
| −3.0 ≤ b < −2.5 | 0 | 0 | 0 | 0 |
| −2.5 ≤ b < −2.0 | 0 | 0 | 0 | 0 |
| −2.0 ≤ b < −1.5 | 0 | 1 | 0 | 1 |
| −1.5 ≤ b < −1.0 | 1 | 1 | 0 | 2 |
| −1.0 ≤ b < −0.5 | 1 | 0 | 4 | 5 |
| −0.5 ≤ b < 0 | 6 | 7 | 2 | 15 |
| 0 ≤ b < 0.5 | 6 | 4 | 5 | 15 |
| 0.5 ≤ b < 1.0 | 3 | 5 | 8 | 16 |
| 1.0 ≤ b < 1.5 | 1 | 1 | 2 | 4 |
| 1.5 ≤ b < 2.0 | 0 | 3 | 2 | 5 |
| 2.0 ≤ b < 2.5 | 0 | 1 | 0 | 1 |
| 2.5 ≤ b < 3.0 | 1 | 0 | 0 | 1 |
| 3.0 ≤ b < 3.5 | 2 | 0 | 0 | 2 |
| b ≥ 3.5 | 0 | 0 | 0 | 0 |
| Minimum | −1.30 | −1.92 | −0.76 | −1.92 |
| Maximum | 3.16 | 2.26 | 1.90 | 3.16 |
| Mean | 0.50 | 0.42 | 0.46 | 0.46 |
| SD | 1.16 | 0.96 | 0.73 | 0.94 |
| **Number of Items:** | **21** | **23** | **23** | **67** |

### Appendix 8.F: Response Time Analyses

**Notes:**

* Response time analyses were based on students who logged on to the assessment and whose total testing time at the test level did not equal zero.
* In table 8.F.1, PT refers to performance task, and HS refers to high school. According to the test design, half of the students received a PT and the other half of the students received a discrete item block in Segment C. Segment C (PT) provides a summary of the testing time to complete Segment C if students received a PT. Segment C (Discrete) provides a summary of the testing time to complete Segment C if students received a discrete item block.
* Note that the criterion for students to be included in table 8.F.1 is that none of their items are designated as “not seen” items.
* Because testing time was recorded at the page level, PT items that were on a page with multiple items were excluded in the analysis in table 8.F.2.
* The following abbreviations apply in table 8.F.2:
* MC = multiple-choice item, which contains both multiple-choice, single-select; and multiple-choice, multiple-select (MCMS) items.
* CR = constructed-response item.
* TEI = technology-enhanced item. (Note that MCMS items are no longer included in the TEI category and counted toward the MC category.)
* Composite = composite item (an item type that includes multiple parts).
* HS = high school.
* Field test items (Segment C) were excluded from the analysis in table 8.F.3 because students receive PTs from different domains in Segment C.
* The criterion for students to be included in table 8.F.3 is that none of their items in segments A and B are designated as “not seen” items.

Table 8.F.1 Testing Time (in Minutes) by Segment

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Segment** | **N** | **Mean** | **SD** | **Min** | **Max** | **1st Percentile** | **10th Percentile** | **25th Percentile** | **50th Percentile** | **75th Percentile** | **90th Percentile** | **99th Percentile** |
| Grade 5 | Discrete Operational | 421,317 | 70.1 | 43.5 | 0.3 | 892.1 | 13.1 | 31.1 | 42.3 | 59.3 | 85.2 | 121.1 | 231.6 |
| Grade 5 | PT Operational | 421,317 | 35.8 | 24.1 | 0.0 | 740.6 | 3.2 | 12.2 | 20.5 | 30.8 | 45.0 | 63.8 | 121.9 |
| Grade 5 | Discrete Field Test | 210,747 | 13.1 | 10.8 | 0.0 | 298.0 | 1.1 | 4.3 | 6.7 | 10.3 | 16.1 | 24.8 | 54.8 |
| Grade 5 | PT Field Test | 210,570 | 10.3 | 8.6 | 0.0 | 221.8 | 0.7 | 2.8 | 5.1 | 8.3 | 12.8 | 19.4 | 43.1 |
| Grade 8 | Discrete Operational | 424,916 | 65.3 | 38.6 | 0.5 | 932.8 | 7.6 | 27.8 | 40.5 | 57.4 | 81.0 | 111.7 | 199.8 |
| Grade 8 | PT Operational | 424,916 | 22.2 | 15.8 | 0.0 | 504.4 | 1.8 | 6.6 | 12.4 | 19.4 | 28.2 | 39.6 | 77.9 |
| Grade 8 | Discrete Field Test | 212,468 | 10.6 | 8.6 | 0.0 | 264.6 | 0.6 | 2.7 | 5.6 | 8.7 | 13.2 | 19.9 | 43.1 |
| Grade 8 | PT Field Test | 212,448 | 7.0 | 6.3 | 0.0 | 192.5 | 0.4 | 1.5 | 3.2 | 5.6 | 8.8 | 13.2 | 29.8 |
| HS—Grade 10 | Discrete Operational | 24,644 | 52.5 | 26.6 | 0.4 | 549.3 | 6.8 | 23.7 | 35.1 | 49.0 | 65.2 | 84.4 | 138.4 |
| HS—Grade 10 | PT Operational | 24,644 | 18.0 | 12.3 | 0.0 | 202.4 | 1.6 | 5.0 | 9.6 | 16.1 | 23.5 | 32.4 | 60.5 |
| HS—Grade 10 | Discrete Field Test | 12,257 | 7.8 | 6.0 | 0.0 | 124.7 | 0.5 | 2.1 | 4.1 | 6.7 | 10.0 | 14.2 | 30.0 |
| HS—Grade 10 | PT Field Test | 12,387 | 5.2 | 4.3 | 0.0 | 83.1 | 0.4 | 1.2 | 2.5 | 4.5 | 6.8 | 9.6 | 20.5 |
| HS—Grade 11 | Discrete Operational | 329,670 | 48.8 | 28.2 | 0.5 | 719.8 | 4.9 | 18.8 | 31.2 | 44.6 | 60.8 | 81.1 | 143.7 |
| HS—Grade 11 | PT Operational | 329,670 | 18.0 | 13.3 | 0.0 | 646.3 | 1.4 | 4.1 | 9.0 | 16.1 | 23.8 | 32.8 | 63.1 |
| HS—Grade 11 | Discrete Field Test | 164,588 | 7.4 | 6.0 | 0.0 | 216.2 | 0.5 | 1.7 | 3.7 | 6.3 | 9.5 | 13.8 | 28.8 |
| HS—Grade 11 | PT Field Test | 165,082 | 5.2 | 4.7 | 0.0 | 433.2 | 0.4 | 1.0 | 2.5 | 4.5 | 6.8 | 9.6 | 21.0 |
| HS—Grade 12 | Discrete Operational | 98,632 | 41.4 | 24.7 | 1.3 | 721.9 | 4.0 | 14.0 | 25.5 | 38.1 | 52.7 | 70.4 | 122.3 |
| HS—Grade 12 | PT Operational | 98,632 | 14.4 | 11.1 | 0.0 | 386.8 | 1.2 | 2.9 | 6.5 | 12.5 | 19.5 | 27.2 | 51.9 |
| HS—Grade 12 | Discrete Field Test | 49,364 | 6.1 | 5.2 | 0.0 | 125.2 | 0.4 | 1.1 | 2.8 | 5.1 | 8.0 | 11.7 | 24.7 |
| HS—Grade 12 | PT Field Test | 49,268 | 4.2 | 3.6 | 0.0 | 102.4 | 0.3 | 0.7 | 1.7 | 3.6 | 5.7 | 8.1 | 16.5 |
| HS—All Grades | Discrete Operational | 452,946 | 47.4 | 27.6 | 0.4 | 721.9 | 4.7 | 17.7 | 30.0 | 43.5 | 59.4 | 79.2 | 139.5 |
| HS—All Grades | PT Operational | 452,946 | 17.3 | 12.9 | 0.0 | 646.3 | 1.3 | 3.8 | 8.4 | 15.3 | 22.9 | 31.7 | 60.9 |
| HS—All Grades | Discrete Field Test | 226,209 | 7.2 | 5.8 | 0.0 | 216.2 | 0.5 | 1.6 | 3.5 | 6.0 | 9.2 | 13.4 | 28.0 |
| HS—All Grades | PT Field Test | 226,737 | 5.0 | 4.5 | 0.0 | 433.2 | 0.4 | 0.9 | 2.3 | 4.3 | 6.5 | 9.3 | 20.1 |

Table 8.F.2 Testing Time (in Minutes) per Item by Item Type

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Item Type** | **N** | **Mean** | **SD** | **Min** | **Max** | **1st Percentile** | **10th Percentile** | **25th Percentile** | **50th Percentile** | **75th Percentile** | **90th Percentile** | **99th Percentile** |
| Grade 5 | MC | 7,174,905 | 1.9 | 2.5 | 0.0 | 413.9 | 0.1 | 0.4 | 0.7 | 1.2 | 2.2 | 3.8 | 12.0 |
| Grade 5 | CR | 1,371,461 | 8.3 | 8.7 | 0.0 | 401.8 | 0.2 | 1.9 | 3.4 | 5.9 | 10.2 | 16.9 | 42.7 |
| Grade 5 | TEI | 4,432,496 | 2.0 | 2.5 | 0.0 | 161.5 | 0.1 | 0.4 | 0.8 | 1.3 | 2.3 | 4.0 | 12.1 |
| Grade 5 | Composite | 1,372,200 | 2.5 | 2.8 | 0.0 | 329.4 | 0.3 | 0.8 | 1.1 | 1.7 | 2.8 | 4.6 | 13.3 |
| Grade 8 | MC | 6,507,501 | 1.8 | 2.5 | 0.0 | 137.0 | 0.0 | 0.1 | 0.6 | 1.2 | 2.1 | 3.7 | 11.7 |
| Grade 8 | CR | 1,630,681 | 5.5 | 6.2 | 0.0 | 410.4 | 0.1 | 1.1 | 2.1 | 3.8 | 6.6 | 11.2 | 30.7 |
| Grade 8 | TEI | 6,376,009 | 1.7 | 2.2 | 0.0 | 151.9 | 0.1 | 0.2 | 0.6 | 1.1 | 2.0 | 3.4 | 10.2 |
| Grade 8 | Composite | 480,126 | 2.8 | 2.7 | 0.0 | 130.0 | 0.2 | 0.7 | 1.5 | 2.2 | 3.3 | 5.0 | 13.2 |
| HS—Grade 10 | MC | 416,666 | 1.4 | 1.7 | 0.0 | 111.6 | 0.0 | 0.1 | 0.5 | 1.0 | 1.8 | 2.9 | 7.6 |
| HS—Grade 10 | CR | 93,135 | 3.8 | 4.0 | 0.0 | 137.0 | 0.1 | 0.7 | 1.6 | 2.9 | 4.8 | 7.7 | 19.0 |
| HS—Grade 10 | TEI | 372,793 | 1.2 | 1.5 | 0.0 | 81.8 | 0.1 | 0.2 | 0.5 | 0.9 | 1.5 | 2.5 | 6.6 |
| HS—Grade 10 | Composite | 49,754 | 1.8 | 1.8 | 0.0 | 83.8 | 0.1 | 0.6 | 0.9 | 1.4 | 2.1 | 3.2 | 7.8 |
| HS—Grade 11 | MC | 5,550,939 | 1.3 | 1.7 | 0.0 | 132.9 | 0.0 | 0.1 | 0.4 | 0.9 | 1.7 | 2.7 | 7.2 |
| HS—Grade 11 | CR | 1,242,649 | 3.6 | 4.0 | 0.0 | 288.5 | 0.1 | 0.6 | 1.5 | 2.7 | 4.5 | 7.3 | 18.7 |
| HS—Grade 11 | TEI | 4,969,514 | 1.2 | 1.4 | 0.0 | 189.0 | 0.1 | 0.2 | 0.4 | 0.8 | 1.4 | 2.3 | 6.2 |
| HS—Grade 11 | Composite | 662,836 | 1.7 | 1.7 | 0.0 | 115.7 | 0.1 | 0.5 | 0.9 | 1.3 | 2.0 | 3.0 | 7.5 |
| HS—Grade 12 | MC | 1,664,928 | 1.1 | 1.4 | 0.0 | 86.3 | 0.0 | 0.1 | 0.3 | 0.8 | 1.4 | 2.3 | 6.0 |
| HS—Grade 12 | CR | 372,489 | 3.0 | 3.4 | 0.0 | 240.0 | 0.1 | 0.4 | 1.1 | 2.2 | 3.9 | 6.3 | 15.7 |
| HS—Grade 12 | TEI | 1,490,354 | 1.0 | 1.3 | 0.0 | 179.5 | 0.1 | 0.2 | 0.4 | 0.7 | 1.2 | 2.0 | 5.4 |
| HS—Grade 12 | Composite | 198,787 | 1.5 | 1.5 | 0.0 | 89.0 | 0.1 | 0.4 | 0.8 | 1.2 | 1.8 | 2.7 | 6.7 |
| HS—All Grades | MC | 7,632,533 | 1.3 | 1.6 | 0.0 | 132.9 | 0.0 | 0.1 | 0.4 | 0.9 | 1.6 | 2.6 | 7.0 |
| HS—All Grades | CR | 1,708,273 | 3.5 | 3.9 | 0.0 | 288.5 | 0.1 | 0.5 | 1.4 | 2.6 | 4.4 | 7.1 | 18.1 |
| HS—All Grades | TEI | 6,832,661 | 1.1 | 1.4 | 0.0 | 189.0 | 0.1 | 0.2 | 0.4 | 0.8 | 1.4 | 2.3 | 6.0 |
| HS—All Grades | Composite | 911,377 | 1.6 | 1.7 | 0.0 | 115.7 | 0.1 | 0.5 | 0.9 | 1.3 | 2.0 | 2.9 | 7.4 |

Table 8.F. Testing Time (in Minutes) for Operational Items by Content Domain

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Content Domain** | **N** | **Mean** | **SD** | **Min** | **Max** | **1st Percentile** | **10th Percentile** | **25th Percentile** | **50th Percentile** | **75th Percentile** | **90th Percentile** | **99th Percentile** |
| Grade 5 | ESS | 421,332 | 37.7 | 23.8 | 0.3 | 532.2 | 6.1 | 15.9 | 22.5 | 32.1 | 46.4 | 65.7 | 124.9 |
| Grade 5 | LS | 421,332 | 37.6 | 23.5 | 0.0 | 403.0 | 5.6 | 16.1 | 22.6 | 32.0 | 46.0 | 65.2 | 123.3 |
| Grade 5 | PS | 421,332 | 30.5 | 19.6 | 0.0 | 520.2 | 5.0 | 12.8 | 18.1 | 25.8 | 37.3 | 53.2 | 103.1 |
| Grade 8 | ESS | 424,941 | 23.6 | 15.0 | 0.0 | 336.6 | 2.2 | 9.1 | 14.3 | 20.6 | 29.4 | 41.0 | 77.2 |
| Grade 8 | LS | 424,941 | 35.3 | 21.1 | 0.1 | 460.9 | 4.1 | 14.7 | 21.9 | 31.1 | 43.7 | 60.1 | 110.2 |
| Grade 8 | PS | 424,941 | 28.6 | 17.7 | 0.0 | 463.7 | 3.1 | 11.5 | 17.5 | 25.1 | 35.5 | 49.1 | 91.5 |
| HS—Grade 10 | ESS | 24,648 | 26.1 | 13.7 | 0.0 | 238.5 | 3.3 | 11.2 | 17.4 | 24.3 | 32.4 | 42.2 | 70.3 |
| HS—Grade 10 | LS | 24,648 | 23.5 | 13.0 | 0.0 | 327.4 | 2.6 | 9.4 | 15.1 | 21.7 | 29.4 | 38.9 | 66.4 |
| HS—Grade 10 | PS | 24,648 | 21.0 | 11.5 | 0.0 | 216.1 | 2.7 | 9.0 | 13.6 | 19.3 | 26.0 | 34.4 | 59.4 |
| HS—Grade 11 | ESS | 329,696 | 24.7 | 14.6 | 0.0 | 445.7 | 2.6 | 9.0 | 15.6 | 22.8 | 30.9 | 41.3 | 73.7 |
| HS—Grade 11 | LS | 329,696 | 22.3 | 13.9 | 0.1 | 408.0 | 1.8 | 7.5 | 13.6 | 20.4 | 28.2 | 38.0 | 69.4 |
| HS—Grade 11 | PS | 329,696 | 19.8 | 12.2 | 0.1 | 477.1 | 2.0 | 7.2 | 12.2 | 17.9 | 24.7 | 33.3 | 61.4 |
| HS—Grade 12 | ESS | 98,641 | 20.8 | 12.7 | 0.0 | 258.6 | 2.1 | 6.8 | 12.5 | 19.2 | 26.5 | 35.5 | 62.6 |
| HS—Grade 12 | LS | 98,641 | 18.5 | 11.9 | 0.5 | 411.8 | 1.5 | 5.4 | 10.6 | 16.9 | 23.9 | 32.3 | 58.2 |
| HS—Grade 12 | PS | 98,641 | 16.5 | 10.5 | 0.4 | 410.6 | 1.6 | 5.4 | 9.7 | 15.0 | 21.0 | 28.4 | 51.6 |
| HS—All Grades | ESS | 452,985 | 24.0 | 14.3 | 0.0 | 445.7 | 2.4 | 8.5 | 14.9 | 22.1 | 30.1 | 40.2 | 71.6 |
| HS—All Grades | LS | 452,985 | 21.6 | 13.5 | 0.0 | 411.8 | 1.8 | 7.0 | 12.9 | 19.7 | 27.4 | 36.9 | 67.1 |
| HS—All Grades | PS | 452,985 | 19.1 | 11.9 | 0.0 | 477.1 | 1.9 | 6.8 | 11.7 | 17.3 | 24.0 | 32.4 | 59.6 |

### Appendix 8.G: Reliability Analyses

Table 8.G.1 Reliabilities and SEMs by Demographic Student Groups for Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Reliability** | **Theta Score SEM** | **Theta Score Variance** |
| All students | 421,426 | 0.92 | 0.31 | 1.13 |
| Male | 214,994 | 0.92 | 0.31 | 1.19 |
| Female | 206,316 | 0.91 | 0.30 | 1.06 |
| Nonbinary | 116 | 0.92 | 0.32 | 1.37 |
| EL | 80,656 | 0.81 | 0.30 | 0.47 |
| English only | 258,971 | 0.91 | 0.31 | 1.12 |
| RFEP | 55,364 | 0.89 | 0.30 | 0.83 |
| IFEP | 26,303 | 0.90 | 0.33 | 1.08 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 69 | 0.84 | 0.35 | 0.74 |
| Economically disadvantaged | 277,118 | 0.90 | 0.30 | 0.91 |
| Not economically disadvantaged | 144,308 | 0.90 | 0.33 | 1.14 |
| American Indian or Alaska Native (All) | 1,806 | 0.90 | 0.30 | 0.95 |
| Asian (All) | 44,552 | 0.90 | 0.35 | 1.28 |
| Native Hawaiian or Other Pacific Islander (All) | 1,708 | 0.90 | 0.29 | 0.89 |
| Filipino (All) | 9,378 | 0.89 | 0.32 | 0.93 |
| Hispanic or Latino (All) | 234,848 | 0.90 | 0.29 | 0.85 |
| Black or African American (All) | 20,869 | 0.90 | 0.30 | 0.87 |
| White (All) | 84,133 | 0.91 | 0.32 | 1.11 |
| Two or more races (All) | 24,132 | 0.91 | 0.33 | 1.24 |
| Disability | 56,535 | 0.89 | 0.31 | 0.90 |
| No disability | 364,891 | 0.91 | 0.31 | 1.07 |
| Migrant education | 3,438 | 0.88 | 0.30 | 0.71 |
| Not migrant education | 417,988 | 0.92 | 0.31 | 1.13 |
| Armed forces family member | 6,480 | 0.91 | 0.30 | 1.00 |
| Not armed forces family member | 414,946 | 0.92 | 0.31 | 1.13 |
| Homeless | 15,562 | 0.89 | 0.30 | 0.83 |
| Not homeless | 405,864 | 0.92 | 0.31 | 1.13 |
| Foster youth | 1,355 | 0.88 | 0.30 | 0.73 |
| Not foster youth | 420,071 | 0.92 | 0.31 | 1.13 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 465 | 0.91 | 0.31 | 1.13 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 1,341 | 0.89 | 0.30 | 0.78 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 25,349 | 0.88 | 0.37 | 1.12 |
| Asian (Primary ethnicity—Economically disadvantaged) | 19,203 | 0.91 | 0.32 | 1.22 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 425 | 0.91 | 0.30 | 0.96 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 1,283 | 0.90 | 0.29 | 0.82 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 5,287 | 0.88 | 0.33 | 0.91 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 4,091 | 0.90 | 0.31 | 0.89 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 45,388 | 0.91 | 0.30 | 1.00 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 189,460 | 0.89 | 0.29 | 0.78 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 4,131 | 0.91 | 0.30 | 0.99 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 16,738 | 0.88 | 0.30 | 0.77 |
| White (Primary ethnicity—Not economically disadvantaged) | 50,058 | 0.89 | 0.33 | 1.00 |
| White (Primary ethnicity—Economically disadvantaged) | 34,075 | 0.91 | 0.30 | 1.05 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 13,205 | 0.89 | 0.34 | 1.11 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 10,927 | 0.91 | 0.30 | 1.07 |

Table 8.G.2 Reliabilities and SEMs by Demographic Student Groups for Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Reliability** | **Theta Score SEM** | **Theta Score Variance** |
| All students | 425,735 | 0.91 | 0.31 | 1.13 |
| Male | 217,872 | 0.92 | 0.32 | 1.22 |
| Female | 207,484 | 0.91 | 0.31 | 1.04 |
| Nonbinary | 379 | 0.91 | 0.31 | 1.04 |
| EL | 54,362 | 0.69 | 0.34 | 0.37 |
| English only | 242,097 | 0.92 | 0.31 | 1.15 |
| RFEP | 111,276 | 0.90 | 0.30 | 0.89 |
| IFEP | 17,907 | 0.91 | 0.34 | 1.26 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 57 | 0.79 | 0.36 | 0.62 |
| Economically disadvantaged | 277,900 | 0.89 | 0.31 | 0.88 |
| Not economically disadvantaged | 147,835 | 0.92 | 0.32 | 1.23 |
| American Indian or Alaska Native (All) | 1,804 | 0.89 | 0.31 | 0.88 |
| Asian (All) | 41,729 | 0.91 | 0.34 | 1.31 |
| Native Hawaiian or Other Pacific Islander (All) | 1,847 | 0.90 | 0.31 | 0.91 |
| Filipino (All) | 10,027 | 0.90 | 0.31 | 0.94 |
| Hispanic or Latino (All) | 242,291 | 0.89 | 0.30 | 0.83 |
| Black or African American (All) | 21,433 | 0.87 | 0.32 | 0.79 |
| White (All) | 84,957 | 0.92 | 0.31 | 1.16 |
| Two or more races (All) | 21,647 | 0.92 | 0.32 | 1.26 |
| Disability | 50,619 | 0.84 | 0.34 | 0.70 |
| No disability | 375,116 | 0.91 | 0.31 | 1.11 |
| Migrant education | 3,341 | 0.86 | 0.32 | 0.72 |
| Not migrant education | 422,394 | 0.91 | 0.31 | 1.13 |
| Armed forces family member | 5,964 | 0.91 | 0.31 | 1.07 |
| Not armed forces family member | 419,771 | 0.91 | 0.31 | 1.13 |
| Homeless | 13,770 | 0.87 | 0.32 | 0.78 |
| Not homeless | 411,965 | 0.91 | 0.31 | 1.14 |
| Foster youth | 1,431 | 0.84 | 0.33 | 0.69 |
| Not foster youth | 424,304 | 0.91 | 0.31 | 1.13 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 525 | 0.90 | 0.29 | 0.88 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 1,279 | 0.87 | 0.32 | 0.81 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 23,981 | 0.89 | 0.36 | 1.16 |
| Asian (Primary ethnicity—Economically disadvantaged) | 17,748 | 0.92 | 0.32 | 1.24 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 514 | 0.91 | 0.31 | 1.08 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 1,333 | 0.87 | 0.31 | 0.74 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 5,898 | 0.90 | 0.31 | 0.95 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 4,129 | 0.90 | 0.29 | 0.87 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 48,421 | 0.91 | 0.30 | 0.99 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 193,870 | 0.88 | 0.31 | 0.75 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 4,890 | 0.90 | 0.30 | 0.91 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 16,543 | 0.86 | 0.32 | 0.71 |
| White (Primary ethnicity—Not economically disadvantaged) | 51,376 | 0.91 | 0.32 | 1.11 |
| White (Primary ethnicity—Economically disadvantaged) | 33,581 | 0.91 | 0.30 | 1.02 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 12,230 | 0.91 | 0.33 | 1.20 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 9,417 | 0.91 | 0.31 | 1.02 |

Table 8.G.3 Reliabilities and SEMs by Demographic Student Groups for Grade Ten

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Reliability** | **Theta Score SEM** | **Theta Score Variance** |
| All students | 24,907 | 0.91 | 0.28 | 0.90 |
| Male | 12,791 | 0.92 | 0.29 | 0.97 |
| Female | 12,052 | 0.90 | 0.28 | 0.82 |
| Nonbinary | 64 | 0.92 | 0.28 | 0.98 |
| EL | 3,070 | 0.69 | 0.30 | 0.29 |
| English only | 13,742 | 0.91 | 0.28 | 0.92 |
| RFEP | 7,069 | 0.90 | 0.27 | 0.74 |
| IFEP | 980 | 0.92 | 0.30 | 1.16 |
| ADEL | 41 | 0.81 | 0.30 | 0.46 |
| To be determined | 0 | N/A | N/A | N/A |
| Economically disadvantaged | 16,710 | 0.89 | 0.28 | 0.72 |
| Not economically disadvantaged | 8,197 | 0.92 | 0.29 | 1.07 |
| American Indian or Alaska Native (All) | 129 | 0.92 | 0.29 | 1.02 |
| Asian (All) | 1,622 | 0.91 | 0.31 | 1.14 |
| Native Hawaiian or Other Pacific Islander (All) | 84 | 0.88 | 0.28 | 0.62 |
| Filipino (All) | 559 | 0.90 | 0.28 | 0.79 |
| Hispanic or Latino (All) | 15,377 | 0.88 | 0.28 | 0.66 |
| Black or African American (All) | 1,463 | 0.86 | 0.29 | 0.59 |
| White (All) | 4,667 | 0.92 | 0.28 | 0.94 |
| Two or more races (All) | 1,006 | 0.93 | 0.29 | 1.15 |
| Disability | 3,089 | 0.82 | 0.30 | 0.50 |
| No disability | 21,818 | 0.91 | 0.28 | 0.89 |
| Migrant education | 233 | 0.87 | 0.28 | 0.61 |
| Not migrant education | 24,674 | 0.91 | 0.28 | 0.90 |
| Armed forces family member | 313 | 0.91 | 0.28 | 0.88 |
| Not armed forces family member | 24,594 | 0.91 | 0.28 | 0.90 |
| Homeless | 758 | 0.88 | 0.28 | 0.66 |
| Not homeless | 24,149 | 0.91 | 0.28 | 0.90 |
| Foster youth | 150 | 0.88 | 0.30 | 0.77 |
| Not foster youth | 24,757 | 0.91 | 0.28 | 0.90 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 39 | 0.92 | 0.28 | 0.96 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 90 | 0.90 | 0.30 | 0.91 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 876 | 0.90 | 0.32 | 1.09 |
| Asian (Primary ethnicity—Economically disadvantaged) | 746 | 0.92 | 0.30 | 1.08 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 35 | 0.90 | 0.28 | 0.78 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 49 | 0.78 | 0.28 | 0.35 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 351 | 0.90 | 0.28 | 0.79 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 208 | 0.90 | 0.27 | 0.76 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 3,219 | 0.90 | 0.28 | 0.79 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 12,158 | 0.87 | 0.28 | 0.61 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 364 | 0.89 | 0.28 | 0.75 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 1,099 | 0.84 | 0.29 | 0.51 |
| White (Primary ethnicity—Not economically disadvantaged) | 2,754 | 0.92 | 0.29 | 1.00 |
| White (Primary ethnicity—Economically disadvantaged) | 1,913 | 0.90 | 0.28 | 0.76 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 559 | 0.92 | 0.29 | 1.16 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 447 | 0.91 | 0.29 | 0.91 |

Table 8.G.4 Reliabilities and SEMs by Demographic Student Groups for Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Reliability** | **Theta Score SEM** | **Theta Score Variance** |
| All students | 331,315 | 0.92 | 0.29 | 1.07 |
| Male | 169,491 | 0.93 | 0.29 | 1.18 |
| Female | 161,153 | 0.92 | 0.28 | 0.96 |
| Nonbinary | 671 | 0.92 | 0.29 | 1.03 |
| EL | 35,104 | 0.71 | 0.30 | 0.32 |
| English only | 178,454 | 0.92 | 0.29 | 1.11 |
| RFEP | 102,137 | 0.91 | 0.28 | 0.85 |
| IFEP | 15,569 | 0.92 | 0.30 | 1.21 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 26 | 0.83 | 0.34 | 0.70 |
| Economically disadvantaged | 206,549 | 0.91 | 0.28 | 0.85 |
| Not economically disadvantaged | 124,766 | 0.93 | 0.30 | 1.20 |
| American Indian or Alaska Native (All) | 1,426 | 0.91 | 0.29 | 0.92 |
| Asian (All) | 33,042 | 0.92 | 0.31 | 1.22 |
| Native Hawaiian or Other Pacific Islander (All) | 1,410 | 0.91 | 0.28 | 0.90 |
| Filipino (All) | 8,613 | 0.91 | 0.29 | 0.90 |
| Hispanic or Latino (All) | 187,637 | 0.90 | 0.28 | 0.80 |
| Black or African American (All) | 14,690 | 0.89 | 0.29 | 0.76 |
| White (All) | 69,376 | 0.93 | 0.29 | 1.15 |
| Two or more races (All) | 15,121 | 0.93 | 0.30 | 1.25 |
| Disability | 35,584 | 0.84 | 0.30 | 0.57 |
| No disability | 295,731 | 0.92 | 0.29 | 1.05 |
| Migrant education | 2,252 | 0.88 | 0.29 | 0.72 |
| Not migrant education | 329,063 | 0.92 | 0.29 | 1.07 |
| Armed forces family member | 4,968 | 0.92 | 0.28 | 0.98 |
| Not armed forces family member | 326,347 | 0.92 | 0.29 | 1.07 |
| Homeless | 10,108 | 0.88 | 0.29 | 0.73 |
| Not homeless | 321,207 | 0.92 | 0.29 | 1.08 |
| Foster youth | 1,133 | 0.87 | 0.30 | 0.69 |
| Not foster youth | 330,182 | 0.92 | 0.29 | 1.07 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 502 | 0.92 | 0.29 | 1.02 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 924 | 0.89 | 0.30 | 0.79 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 18,529 | 0.91 | 0.32 | 1.13 |
| Asian (Primary ethnicity—Economically disadvantaged) | 14,513 | 0.92 | 0.29 | 1.12 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 507 | 0.92 | 0.28 | 1.00 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 903 | 0.90 | 0.28 | 0.80 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 5,173 | 0.91 | 0.29 | 0.91 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 3,440 | 0.91 | 0.28 | 0.84 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 41,313 | 0.92 | 0.28 | 0.96 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 146,324 | 0.89 | 0.28 | 0.73 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 4,194 | 0.91 | 0.28 | 0.90 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 10,496 | 0.87 | 0.29 | 0.66 |
| White (Primary ethnicity—Not economically disadvantaged) | 45,416 | 0.92 | 0.30 | 1.14 |
| White (Primary ethnicity—Economically disadvantaged) | 23,960 | 0.92 | 0.28 | 0.99 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 9,132 | 0.92 | 0.30 | 1.21 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 5,989 | 0.92 | 0.29 | 1.05 |

Table 8.G.5 Reliabilities and SEMs by Demographic Student Groups for Grade Twelve

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Reliability** | **Theta Score SEM** | **Theta Score Variance** |
| All students | 99,413 | 0.92 | 0.29 | 1.09 |
| Male | 50,813 | 0.93 | 0.30 | 1.19 |
| Female | 48,388 | 0.92 | 0.29 | 0.98 |
| Nonbinary | 212 | 0.91 | 0.28 | 0.91 |
| EL | 12,020 | 0.69 | 0.31 | 0.31 |
| English only | 50,771 | 0.93 | 0.29 | 1.15 |
| RFEP | 31,035 | 0.91 | 0.28 | 0.88 |
| IFEP | 5,340 | 0.93 | 0.30 | 1.25 |
| ADEL | 207 | 0.75 | 0.29 | 0.33 |
| To be determined | 8 | N/A | N/A | N/A |
| Economically disadvantaged | 61,685 | 0.90 | 0.29 | 0.82 |
| Not economically disadvantaged | 37,728 | 0.93 | 0.30 | 1.28 |
| American Indian or Alaska Native (All) | 453 | 0.90 | 0.29 | 0.85 |
| Asian (All) | 11,052 | 0.93 | 0.31 | 1.35 |
| Native Hawaiian or Other Pacific Islander (All) | 530 | 0.89 | 0.29 | 0.73 |
| Filipino (All) | 2,876 | 0.91 | 0.29 | 0.96 |
| Hispanic or Latino (All) | 56,033 | 0.89 | 0.29 | 0.76 |
| Black or African American (All) | 5,528 | 0.88 | 0.29 | 0.69 |
| White (All) | 18,468 | 0.93 | 0.30 | 1.24 |
| Two or more races (All) | 4,473 | 0.93 | 0.30 | 1.27 |
| Disability | 10,912 | 0.85 | 0.30 | 0.60 |
| No disability | 88,501 | 0.92 | 0.29 | 1.09 |
| Migrant education | 818 | 0.85 | 0.29 | 0.56 |
| Not migrant education | 98,595 | 0.92 | 0.29 | 1.09 |
| Armed forces family member | 1,402 | 0.92 | 0.29 | 1.06 |
| Not armed forces family member | 98,011 | 0.92 | 0.29 | 1.09 |
| Homeless | 3,965 | 0.88 | 0.29 | 0.70 |
| Not homeless | 95,448 | 0.92 | 0.29 | 1.10 |
| Foster youth | 506 | 0.82 | 0.31 | 0.54 |
| Not foster youth | 98,907 | 0.92 | 0.29 | 1.09 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 144 | 0.92 | 0.29 | 1.03 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 309 | 0.89 | 0.29 | 0.74 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 6,558 | 0.92 | 0.32 | 1.25 |
| Asian (Primary ethnicity—Economically disadvantaged) | 4,494 | 0.93 | 0.30 | 1.23 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 187 | 0.91 | 0.28 | 0.88 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 343 | 0.86 | 0.29 | 0.62 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 1,820 | 0.91 | 0.29 | 0.92 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 1,056 | 0.92 | 0.29 | 0.99 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 12,422 | 0.91 | 0.29 | 0.94 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 43,611 | 0.88 | 0.29 | 0.69 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 1,628 | 0.90 | 0.29 | 0.86 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 3,900 | 0.85 | 0.30 | 0.59 |
| White (Primary ethnicity—Not economically disadvantaged) | 12,264 | 0.93 | 0.30 | 1.26 |
| White (Primary ethnicity—Economically disadvantaged) | 6,204 | 0.92 | 0.29 | 1.03 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 2,705 | 0.93 | 0.30 | 1.30 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 1,768 | 0.91 | 0.29 | 0.95 |

Table 8.G.6 Reliabilities and SEMs by Demographic Student Groups for High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student Group** | **Number Tested** | **Reliability** | **Theta Score SEM** | **Theta Score Variance** |
| All students | 455,635 | 0.92 | 0.29 | 1.07 |
| Male | 233,095 | 0.93 | 0.29 | 1.17 |
| Female | 221,593 | 0.92 | 0.28 | 0.96 |
| Nonbinary | 947 | 0.92 | 0.29 | 1.00 |
| EL | 50,194 | 0.71 | 0.30 | 0.32 |
| English only | 242,967 | 0.92 | 0.29 | 1.11 |
| RFEP | 140,241 | 0.91 | 0.28 | 0.86 |
| IFEP | 21,889 | 0.92 | 0.30 | 1.22 |
| ADEL | 248 | 0.76 | 0.29 | 0.35 |
| To be determined | 34 | 0.86 | 0.33 | 0.78 |
| Economically disadvantaged | 284,944 | 0.90 | 0.28 | 0.84 |
| Not economically disadvantaged | 170,691 | 0.93 | 0.30 | 1.22 |
| American Indian or Alaska Native (All) | 2,008 | 0.91 | 0.29 | 0.91 |
| Asian (All) | 45,716 | 0.92 | 0.31 | 1.25 |
| Native Hawaiian or Other Pacific Islander (All) | 2,024 | 0.90 | 0.28 | 0.85 |
| Filipino (All) | 12,048 | 0.91 | 0.29 | 0.92 |
| Hispanic or Latino (All) | 259,047 | 0.90 | 0.28 | 0.78 |
| Black or African American (All) | 21,681 | 0.89 | 0.29 | 0.74 |
| White (All) | 92,511 | 0.93 | 0.29 | 1.16 |
| Two or more races (All) | 20,600 | 0.93 | 0.30 | 1.25 |
| Disability | 49,585 | 0.84 | 0.30 | 0.57 |
| No disability | 406,050 | 0.92 | 0.29 | 1.06 |
| Migrant education | 3,303 | 0.88 | 0.29 | 0.68 |
| Not migrant education | 452,332 | 0.92 | 0.29 | 1.07 |
| Armed forces family member | 6,683 | 0.92 | 0.28 | 1.00 |
| Not armed forces family member | 448,952 | 0.92 | 0.29 | 1.07 |
| Homeless | 14,831 | 0.88 | 0.29 | 0.72 |
| Not homeless | 440,804 | 0.92 | 0.29 | 1.08 |
| Foster youth | 1,789 | 0.86 | 0.30 | 0.66 |
| Not foster youth | 453,846 | 0.92 | 0.29 | 1.07 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 685 | 0.92 | 0.29 | 1.02 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 1,323 | 0.89 | 0.29 | 0.79 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 25,963 | 0.91 | 0.32 | 1.16 |
| Asian (Primary ethnicity—Economically disadvantaged) | 19,753 | 0.92 | 0.29 | 1.15 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 729 | 0.92 | 0.28 | 0.97 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 1,295 | 0.89 | 0.29 | 0.74 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 7,344 | 0.91 | 0.29 | 0.91 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 4,704 | 0.91 | 0.28 | 0.88 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 56,954 | 0.92 | 0.28 | 0.95 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 202,093 | 0.89 | 0.28 | 0.72 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 6,186 | 0.91 | 0.28 | 0.89 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 15,495 | 0.87 | 0.29 | 0.64 |
| White (Primary ethnicity—Not economically disadvantaged) | 60,434 | 0.92 | 0.30 | 1.16 |
| White (Primary ethnicity—Economically disadvantaged) | 32,077 | 0.92 | 0.29 | 0.98 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 12,396 | 0.93 | 0.30 | 1.23 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 8,204 | 0.92 | 0.29 | 1.02 |

**Note:** “Total Matched Cases” indicates students who took both CAST and the ELPAC in 2023–24. “Beginning to Develop,” “Somewhat Developed,” “Moderately Developed,” and “Well Developed” indicate the overall English language proficiency level attained by students, as defined by the ELPAC.

Table 8.G.7 Reliabilities and SEMs by ELPAC Performance Levels

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Group** | **Number Tested** | **Reliability** | **Theta Score SEM** | **Theta Score Variance** |
| Grade 5 | Total Matched Cases | 82,417 | 0.82 | 0.29 | 0.47 |
| Grade 5 | Beginning to Develop | 14,529 | 0.59 | 0.34 | 0.29 |
| Grade 5 | Somewhat Developed | 21,479 | 0.61 | 0.31 | 0.24 |
| Grade 5 | Moderately Developed | 29,007 | 0.73 | 0.28 | 0.28 |
| Grade 5 | Well Developed | 17,402 | 0.81 | 0.27 | 0.37 |
| Grade 8 | Total Matched Cases | 55,522 | 0.69 | 0.34 | 0.37 |
| Grade 8 | Beginning to Develop | 11,605 | 0.44 | 0.38 | 0.25 |
| Grade 8 | Somewhat Developed | 12,479 | 0.50 | 0.36 | 0.26 |
| Grade 8 | Moderately Developed | 19,157 | 0.65 | 0.32 | 0.30 |
| Grade 8 | Well Developed | 12,281 | 0.78 | 0.29 | 0.39 |
| High school—Grade 10 | Total Matched Cases | 3,072 | 0.69 | 0.30 | 0.29 |
| High school—Grade 10 | Beginning to Develop | 693 | 0.49 | 0.33 | 0.21 |
| High school—Grade 10 | Somewhat Developed | 767 | 0.56 | 0.31 | 0.21 |
| High school—Grade 10 | Moderately Developed | 1,094 | 0.65 | 0.29 | 0.24 |
| High school—Grade 10 | Well Developed | 518 | 0.75 | 0.27 | 0.30 |
| High school—Grade 11 | Total Matched Cases | 34,992 | 0.72 | 0.30 | 0.32 |
| High school—Grade 11 | Beginning to Develop | 9,699 | 0.48 | 0.33 | 0.21 |
| High school—Grade 11 | Somewhat Developed | 9,566 | 0.58 | 0.30 | 0.22 |
| High school—Grade 11 | Moderately Developed | 10,638 | 0.70 | 0.29 | 0.27 |
| High school—Grade 11 | Well Developed | 5,089 | 0.81 | 0.27 | 0.40 |
| High school—Grade 12 | Total Matched Cases | 11,928 | 0.70 | 0.31 | 0.31 |
| High school—Grade 12 | Beginning to Develop | 3,860 | 0.48 | 0.33 | 0.22 |
| High school—Grade 12 | Somewhat Developed | 3,175 | 0.58 | 0.31 | 0.22 |
| High school—Grade 12 | Moderately Developed | 3,161 | 0.69 | 0.29 | 0.27 |
| High school—Grade 12 | Well Developed | 1,732 | 0.79 | 0.28 | 0.36 |
| High school—All grades | Total Matched Cases | 49,992 | 0.71 | 0.30 | 0.32 |
| High school—All grades | Beginning to Develop | 14,252 | 0.48 | 0.33 | 0.21 |
| High school—All grades | Somewhat Developed | 13,508 | 0.58 | 0.30 | 0.22 |
| High school—All grades | Moderately Developed | 14,893 | 0.69 | 0.29 | 0.27 |
| High school—All grades | Well Developed | 7,339 | 0.80 | 0.27 | 0.38 |

### Appendix 8.H: Conditional Standard Errors of Measurement

CSEM is a measure of the amount of error associated with each theta score. Figure 8.H.1 through figure 8.H.6 present average CSEMs by average theta scores. Data used to generate these figures is presented in table 8.H.1 through table 8.H.6.

#### CSEMs for Grade Five

Figure 8.H.1 plots the average CSEMs by average theta score for grade five. The data used to create this graph is found in the table that immediately follows, table 8.H.1. The graph’s y-‍axis shows the average theta CSEMs corresponding to the theta values ranging from 0.0 to 1.4 in intervals of 0.2. The x-axis shows the average theta scores from −5 to +5 in intervals of 1.

Vertical lines in the CSEM curve represent the thresholds of *Standard Nearly Met*, *Standard Met*, and *Standard Exceeded* achievement levels. In general, for a given theta score, low values of the CSEM indicate greater measurement precision than that provided by larger CSEMs. The average theta scores to reach the *Standard Nearly Met*, *Standard Met*, and *Standard Exceeded* achievement levels are –0.97, 0.51, and 1.35, with average CSEMs of 0.28, 0.27, 0.33, respectively. As expected, the CSEM is highest at the extreme ends of the theta score distribution.

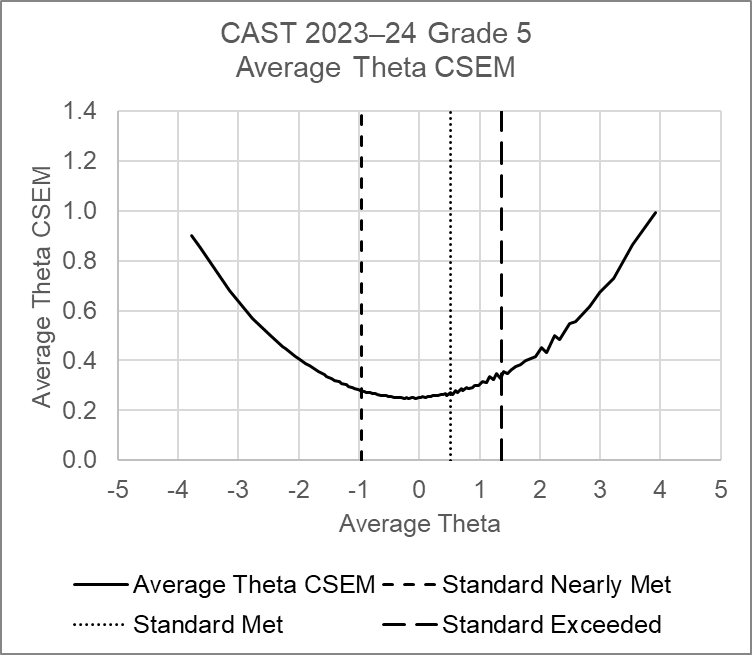


Figure 8.H.1 Average CSEM by theta, grade five

In table 8.H.1, rows where asterisks (\*) precede the values indicate the cut scores. These rows are also shaded.

Table 8.H.1 Average CSEM by Theta Data, Grade Five

|  |  |  |  |
| --- | --- | --- | --- |
| **Scale Score** | **N** | **Average Theta** | **Average Theta CSEM** |
| 153 | 17 | −3.78 | 0.90 |
| 154 | 15 | −3.64 | 0.86 |
| 155 | 137 | −3.15 | 0.68 |
| 157 | 431 | −2.76 | 0.57 |
| 159 | 927 | −2.47 | 0.50 |
| 160 | 430 | −2.27 | 0.46 |
| 161 | 1,392 | −2.22 | 0.45 |
| 162 | 1,580 | −2.06 | 0.42 |
| 163 | 1,598 | −2.00 | 0.41 |
| 164 | 2,288 | −1.88 | 0.39 |
| 165 | 2,436 | −1.83 | 0.38 |
| 166 | 3,023 | −1.72 | 0.36 |
| 167 | 3,257 | −1.67 | 0.36 |
| 168 | 4,063 | −1.58 | 0.34 |
| 169 | 4,208 | −1.53 | 0.34 |
| 170 | 4,743 | −1.45 | 0.33 |
| 171 | 4,890 | −1.40 | 0.32 |
| 172 | 5,295 | −1.33 | 0.32 |
| 173 | 5,575 | −1.29 | 0.31 |
| 174 | 5,695 | −1.21 | 0.30 |
| 175 | 5,893 | −1.18 | 0.30 |
| 176 | 7,362 | −1.10 | 0.29 |
| 177 | 4,578 | −1.07 | 0.29 |
| 178 | 7,570 | −1.00 | 0.29 |
| \*179 | \*4,810 | \*−0.97 | \*0.28 |
| 180 | 9,215 | −0.90 | 0.28 |
| 181 | 3,122 | −0.87 | 0.27 |
| 182 | 9,117 | −0.81 | 0.27 |
| 183 | 3,110 | −0.78 | 0.27 |
| 184 | 9,062 | −0.72 | 0.27 |
| 185 | 3,162 | −0.69 | 0.26 |
| 186 | 10,608 | −0.63 | 0.26 |
| 187 | 3,033 | −0.59 | 0.26 |
| 188 | 8,943 | −0.54 | 0.26 |
| 189 | 2,972 | −0.50 | 0.26 |
| 190 | 8,994 | −0.46 | 0.25 |
| 191 | 4,386 | −0.41 | 0.25 |
| 192 | 8,916 | −0.37 | 0.25 |
| 193 | 2,844 | −0.32 | 0.25 |
| 194 | 8,884 | −0.29 | 0.25 |
| 195 | 4,383 | −0.24 | 0.25 |
| 196 | 5,919 | −0.21 | 0.25 |
| 197 | 5,889 | −0.17 | 0.25 |
| 198 | 5,837 | −0.13 | 0.25 |
| 199 | 5,806 | −0.09 | 0.25 |
| 200 | 4,314 | −0.06 | 0.25 |
| 201 | 7,169 | −0.02 | 0.25 |
| 202 | 4,359 | 0.02 | 0.25 |
| 203 | 7,411 | 0.06 | 0.25 |
| 204 | 4,376 | 0.10 | 0.25 |
| 205 | 7,358 | 0.14 | 0.26 |
| 206 | 4,349 | 0.18 | 0.25 |
| 207 | 5,750 | 0.22 | 0.26 |
| 208 | 5,881 | 0.26 | 0.26 |
| 209 | 5,744 | 0.30 | 0.26 |
| 210 | 5,771 | 0.34 | 0.26 |
| 211 | 5,837 | 0.39 | 0.26 |
| 212 | 5,839 | 0.43 | 0.27 |
| 213 | 2,761 | 0.46 | 0.26 |
| \*214 | \*8,449 | \*0.51 | \*0.27 |
| 215 | 2,781 | 0.54 | 0.26 |
| 216 | 7,013 | 0.60 | 0.28 |
| 217 | 4,357 | 0.63 | 0.27 |
| 218 | 5,650 | 0.69 | 0.29 |
| 219 | 5,613 | 0.72 | 0.28 |
| 220 | 4,158 | 0.79 | 0.29 |
| 221 | 7,060 | 0.81 | 0.29 |
| 222 | 1,283 | 0.88 | 0.29 |
| 223 | 9,625 | 0.91 | 0.30 |
| 224 | 1,292 | 0.99 | 0.30 |
| 225 | 8,030 | 1.02 | 0.31 |
| 226 | 1,342 | 1.05 | 0.32 |
| 227 | 7,525 | 1.12 | 0.31 |
| 228 | 2,676 | 1.16 | 0.34 |
| 229 | 5,919 | 1.24 | 0.32 |
| 230 | 3,795 | 1.28 | 0.35 |
| \*231 | \*3,394 | \*1.35 | \*0.33 |
| 232 | 4,652 | 1.40 | 0.35 |
| 233 | 3,327 | 1.46 | 0.35 |
| 234 | 4,205 | 1.53 | 0.36 |
| 235 | 4,220 | 1.60 | 0.38 |
| 236 | 3,819 | 1.68 | 0.38 |
| 237 | 3,702 | 1.77 | 0.40 |
| 238 | 2,590 | 1.85 | 0.41 |
| 239 | 3,109 | 1.93 | 0.42 |
| 240 | 3,013 | 2.03 | 0.45 |
| 241 | 2,029 | 2.11 | 0.43 |
| 242 | 2,447 | 2.24 | 0.50 |
| 243 | 2,370 | 2.33 | 0.48 |
| 244 | 1,432 | 2.49 | 0.55 |
| 245 | 2,210 | 2.59 | 0.56 |
| 246 | 1,366 | 2.82 | 0.62 |
| 247 | 1,103 | 2.99 | 0.67 |
| 248 | 850 | 3.22 | 0.73 |
| 249 | 642 | 3.53 | 0.86 |
| 250 | 972 | 3.91 | 0.99 |

#### CSEMs for Grade Eight

Figure 8.H.2 plots the average CSEMs by average theta score for grade eight. The data used to create this graph is found in the table that immediately follows, table 8.H.2. The graph’s y-axis shows the average theta CSEMs corresponding to the theta values ranging from 0.0 to 1.4 in intervals of 0.2. The x-axis shows the average theta scores from −5 to +5 in intervals of 1.

Vertical lines in the CSEM curve represent the thresholds of *Standard Nearly Met*, *Standard Met*, and *Standard Exceeded* achievement levels. In general, for a given theta score, low values of the CSEM indicate greater measurement precision than that provided by larger CSEMs. The average theta scores to reach the *Standard Nearly Met*, *Standard Met*, and *Standard Exceeded* achievement levels are –1.01, 0.57, and 1.48, with average CSEMs of 0.30, 0.25, 0.33, respectively. As expected, the CSEM is highest at the extreme ends of the theta score distribution.

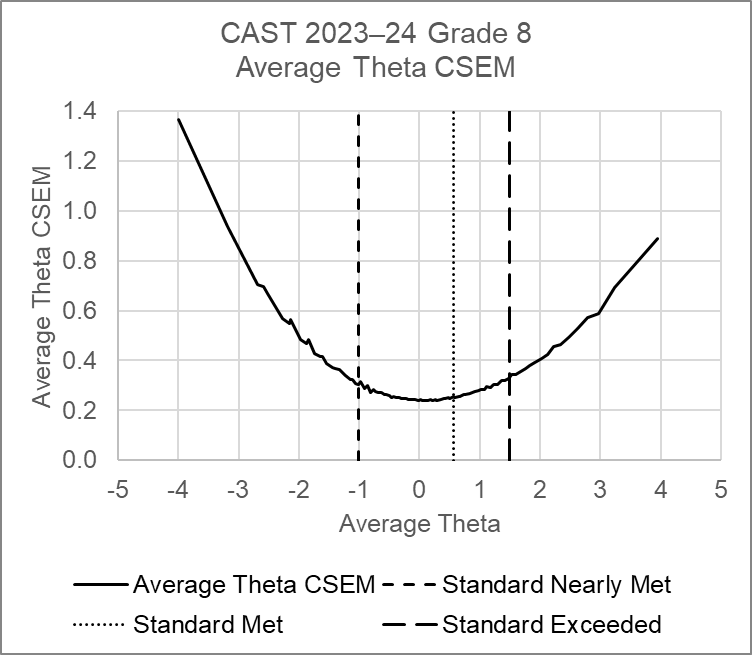


Figure 8.H.2 Average CSEM by theta, grade eight

In table 8.H.2, rows where asterisks (\*) precede the values indicate the cut scores. These rows are also shaded.

Table 8.H.2 Average CSEM by Theta Data, Grade Eight

|  |  |  |  |
| --- | --- | --- | --- |
| **Scale Score** | **N** | **Average Theta** | **Average Theta CSEM** |
| 353 | 82 | −4.00 | 1.37 |
| 355 | 316 | −3.19 | 0.94 |
| 357 | 158 | −2.68 | 0.71 |
| 358 | 780 | −2.59 | 0.70 |
| 359 | 159 | −2.52 | 0.67 |
| 360 | 1,167 | −2.27 | 0.57 |
| 361 | 262 | −2.15 | 0.55 |
| 362 | 1,262 | −2.14 | 0.56 |
| 363 | 2,289 | −1.97 | 0.48 |
| 364 | 586 | −1.87 | 0.47 |
| 365 | 2,408 | −1.85 | 0.48 |
| 366 | 3,736 | −1.74 | 0.43 |
| 367 | 1,005 | −1.64 | 0.42 |
| 368 | 5,036 | −1.60 | 0.42 |
| 369 | 4,125 | −1.54 | 0.39 |
| 370 | 3,035 | −1.45 | 0.37 |
| 371 | 10,166 | −1.39 | 0.37 |
| 372 | 1 | −1.33 | 0.36 |
| 373 | 11,596 | −1.27 | 0.35 |
| 374 | 3,924 | −1.22 | 0.34 |
| 375 | 4,282 | −1.15 | 0.32 |
| 376 | 12,396 | −1.11 | 0.32 |
| 377 | 1 | −1.06 | 0.31 |
| \*378 | \*10,093 | \*−1.01 | \*0.30 |
| 379 | 6,338 | −0.97 | 0.31 |
| 380 | 7,936 | −0.90 | 0.29 |
| 381 | 7,978 | −0.86 | 0.30 |
| 382 | 5,227 | −0.80 | 0.27 |
| 383 | 7,708 | −0.77 | 0.28 |
| 384 | 5,479 | −0.72 | 0.28 |
| 385 | 6,975 | −0.68 | 0.27 |
| 386 | 7,072 | −0.63 | 0.27 |
| 387 | 6,964 | −0.59 | 0.26 |
| 388 | 6,684 | −0.54 | 0.26 |
| 389 | 6,544 | −0.50 | 0.26 |
| 390 | 4,810 | −0.46 | 0.25 |
| 391 | 6,378 | −0.43 | 0.26 |
| 392 | 7,737 | −0.38 | 0.25 |
| 393 | 6,061 | −0.34 | 0.25 |
| 394 | 6,053 | −0.29 | 0.25 |
| 395 | 6,080 | −0.26 | 0.25 |
| 396 | 5,979 | −0.22 | 0.25 |
| 397 | 5,839 | −0.18 | 0.24 |
| 398 | 5,756 | −0.14 | 0.24 |
| 399 | 5,638 | −0.10 | 0.24 |
| 400 | 5,712 | −0.06 | 0.24 |
| 401 | 8,313 | −0.02 | 0.24 |
| 402 | 2,828 | 0.02 | 0.24 |
| 403 | 8,109 | 0.06 | 0.24 |
| 404 | 5,484 | 0.11 | 0.24 |
| 405 | 5,480 | 0.13 | 0.24 |
| 406 | 6,784 | 0.19 | 0.24 |
| 407 | 3,821 | 0.21 | 0.24 |
| 408 | 9,043 | 0.27 | 0.24 |
| 409 | 1,290 | 0.29 | 0.24 |
| 410 | 10,395 | 0.34 | 0.24 |
| 411 | 1,230 | 0.40 | 0.25 |
| 412 | 8,867 | 0.42 | 0.25 |
| 413 | 2,497 | 0.48 | 0.25 |
| 414 | 7,354 | 0.50 | 0.25 |
| \*415 | \*4,732 | \*0.57 | \*0.25 |
| 416 | 4,800 | 0.59 | 0.25 |
| 417 | 4,650 | 0.65 | 0.26 |
| 418 | 4,735 | 0.67 | 0.26 |
| 419 | 6,792 | 0.73 | 0.26 |
| 420 | 2,347 | 0.75 | 0.26 |
| 421 | 8,904 | 0.83 | 0.27 |
| 423 | 8,467 | 0.91 | 0.27 |
| 424 | 1,979 | 0.98 | 0.28 |
| 425 | 6,547 | 1.01 | 0.28 |
| 426 | 2,900 | 1.08 | 0.28 |
| 427 | 4,979 | 1.12 | 0.29 |
| 428 | 2,692 | 1.18 | 0.29 |
| 429 | 4,927 | 1.22 | 0.30 |
| 430 | 4,451 | 1.29 | 0.31 |
| 431 | 3,667 | 1.36 | 0.32 |
| 432 | 3,312 | 1.41 | 0.32 |
| \*433 | \*3,350 | \*1.48 | \*0.33 |
| 434 | 3,141 | 1.53 | 0.34 |
| 435 | 3,113 | 1.60 | 0.34 |
| 436 | 3,676 | 1.68 | 0.36 |
| 437 | 2,882 | 1.76 | 0.37 |
| 438 | 2,694 | 1.83 | 0.38 |
| 439 | 2,578 | 1.91 | 0.39 |
| 440 | 3,557 | 2.02 | 0.41 |
| 441 | 2,265 | 2.13 | 0.42 |
| 442 | 2,073 | 2.23 | 0.45 |
| 443 | 2,389 | 2.34 | 0.46 |
| 444 | 2,158 | 2.49 | 0.50 |
| 445 | 1,523 | 2.63 | 0.53 |
| 446 | 1,777 | 2.79 | 0.57 |
| 447 | 1,382 | 2.98 | 0.59 |
| 448 | 1,440 | 3.24 | 0.69 |
| 449 | 1,066 | 3.60 | 0.79 |
| 450 | 1,713 | 3.94 | 0.89 |

#### CSEMs for Grade Ten

Figure 8.H.3 plots the average CSEMs by average theta score for grade ten. The data used to create this graph is found in the table that immediately follows, table 8.H.3. The graph’s y-‍axis shows the average theta CSEMs corresponding to the theta values ranging from 0.0 to 1.4 in intervals of 0.2. The x-axis shows the average theta scores from −5 to +5 in intervals of 1.

Vertical lines in the CSEM curve represent the thresholds of *Standard Nearly Met*, *Standard Met*, and *Standard Exceeded* achievement levels. In general, for a given theta score, low values of the CSEM indicate greater measurement precision than that provided by larger CSEMs. The average theta scores to reach the *Standard Nearly Met*, *Standard Met*, and *Standard Exceeded* achievement levels are –1.12, 0.56, and 1.67, with average CSEMs of 0.30, 0.25, 0.32, respectively. As expected, the CSEM is highest at the extreme ends of the theta score distribution.

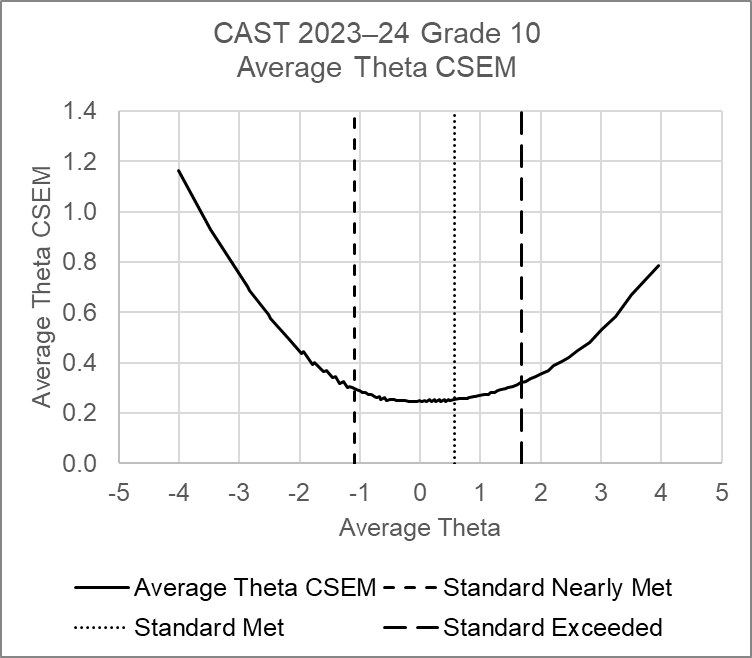


Figure 8.H.3 Average CSEM by theta, grade ten

In table 8.H.3, rows where asterisks (\*) precede the values indicate the cut scores. These rows are also shaded.

Table 8.H.3 Average CSEM by Theta Data, Grade Ten

|  |  |  |  |
| --- | --- | --- | --- |
| **Scale Score** | **N** | **Average Theta** | **Average Theta CSEM** |
| 553 | 1 | −4.00 | 1.16 |
| 554 | 2 | −3.48 | 0.93 |
| 556 | 9 | −2.87 | 0.70 |
| 557 | 1 | −2.82 | 0.69 |
| 558 | 4 | −2.52 | 0.59 |
| 559 | 22 | −2.48 | 0.58 |
| 561 | 54 | −2.18 | 0.50 |
| 563 | 60 | −1.97 | 0.44 |
| 564 | 92 | −1.93 | 0.44 |
| 565 | 22 | −1.79 | 0.39 |
| 566 | 220 | −1.75 | 0.40 |
| 568 | 205 | −1.60 | 0.36 |
| 569 | 146 | −1.55 | 0.37 |
| 570 | 275 | −1.46 | 0.34 |
| 571 | 317 | −1.40 | 0.34 |
| 572 | 299 | −1.33 | 0.32 |
| 573 | 445 | −1.27 | 0.32 |
| 574 | 324 | −1.21 | 0.30 |
| 575 | 311 | −1.17 | 0.31 |
| \*576 | \*480 | \*−1.12 | \*0.30 |
| 577 | 463 | −1.06 | 0.29 |
| 578 | 519 | −1.01 | 0.29 |
| 579 | 506 | −0.96 | 0.28 |
| 580 | 516 | −0.90 | 0.28 |
| 581 | 509 | −0.86 | 0.27 |
| 582 | 521 | −0.81 | 0.27 |
| 583 | 529 | −0.76 | 0.27 |
| 584 | 375 | −0.72 | 0.26 |
| 585 | 490 | −0.68 | 0.27 |
| 586 | 357 | −0.65 | 0.25 |
| 587 | 733 | −0.59 | 0.26 |
| 588 | 216 | −0.56 | 0.25 |
| 589 | 817 | −0.50 | 0.26 |
| 590 | 135 | −0.48 | 0.25 |
| 591 | 756 | −0.42 | 0.25 |
| 592 | 107 | −0.40 | 0.25 |
| 593 | 691 | −0.34 | 0.25 |
| 594 | 95 | −0.31 | 0.25 |
| 595 | 699 | −0.26 | 0.25 |
| 596 | 90 | −0.23 | 0.25 |
| 597 | 661 | −0.18 | 0.25 |
| 598 | 98 | −0.15 | 0.25 |
| 599 | 619 | −0.10 | 0.25 |
| 600 | 279 | −0.05 | 0.25 |
| 601 | 388 | −0.01 | 0.25 |
| 602 | 357 | 0.03 | 0.24 |
| 603 | 364 | 0.07 | 0.25 |
| 604 | 384 | 0.11 | 0.24 |
| 605 | 243 | 0.16 | 0.25 |
| 606 | 380 | 0.18 | 0.24 |
| 607 | 235 | 0.24 | 0.25 |
| 608 | 370 | 0.26 | 0.24 |
| 609 | 247 | 0.32 | 0.25 |
| 610 | 356 | 0.34 | 0.25 |
| 611 | 308 | 0.40 | 0.25 |
| 612 | 253 | 0.41 | 0.25 |
| 613 | 372 | 0.48 | 0.25 |
| 614 | 196 | 0.49 | 0.25 |
| \*615 | \*505 | \*0.56 | \*0.25 |
| 617 | 529 | 0.64 | 0.26 |
| 619 | 487 | 0.72 | 0.26 |
| 620 | 108 | 0.79 | 0.26 |
| 621 | 338 | 0.81 | 0.26 |
| 622 | 324 | 0.88 | 0.26 |
| 623 | 105 | 0.90 | 0.27 |
| 624 | 432 | 0.97 | 0.27 |
| 626 | 415 | 1.06 | 0.27 |
| 627 | 174 | 1.14 | 0.27 |
| 628 | 201 | 1.16 | 0.28 |
| 629 | 229 | 1.24 | 0.28 |
| 630 | 127 | 1.28 | 0.29 |
| 631 | 271 | 1.34 | 0.29 |
| 632 | 150 | 1.42 | 0.30 |
| 633 | 152 | 1.46 | 0.30 |
| 634 | 220 | 1.54 | 0.31 |
| 635 | 133 | 1.62 | 0.31 |
| \*636 | \*152 | \*1.67 | \*0.32 |
| 637 | 158 | 1.76 | 0.33 |
| 638 | 97 | 1.83 | 0.34 |
| 639 | 133 | 1.91 | 0.35 |
| 640 | 117 | 2.02 | 0.36 |
| 641 | 78 | 2.13 | 0.37 |
| 642 | 73 | 2.21 | 0.39 |
| 643 | 115 | 2.34 | 0.40 |
| 644 | 54 | 2.47 | 0.42 |
| 645 | 82 | 2.61 | 0.45 |
| 646 | 38 | 2.81 | 0.48 |
| 647 | 56 | 3.01 | 0.53 |
| 648 | 21 | 3.24 | 0.58 |
| 649 | 26 | 3.51 | 0.67 |
| 650 | 28 | 3.95 | 0.79 |

#### CSEMs for Grade Eleven

Figure 8.H.4 plots the average CSEMs by average theta score for grade eleven. The data used to create this graph is found in the table that immediately follows, table 8.H.4. The graph’s y-axis shows the average theta CSEMs corresponding to the theta values ranging from 0.0 to 1.4 in intervals of 0.2. The x-axis shows the average theta scores from −5 to +5 in intervals of 1.

Vertical lines in the CSEM curve represent the thresholds of *Standard Nearly Met*, *Standard Met*, and *Standard Exceeded* achievement levels. In general, for a given theta score, low values of the CSEM indicate greater measurement precision than that provided by larger CSEMs. The average theta scores to reach the *Standard Nearly Met*, *Standard Met*, and *Standard Exceeded* achievement levels are –1.12, 0.56, and 1.67, with average CSEMs of 0.30, 0.25, 0.32, respectively. As expected, the CSEM is highest at the extreme ends of the theta score distribution.

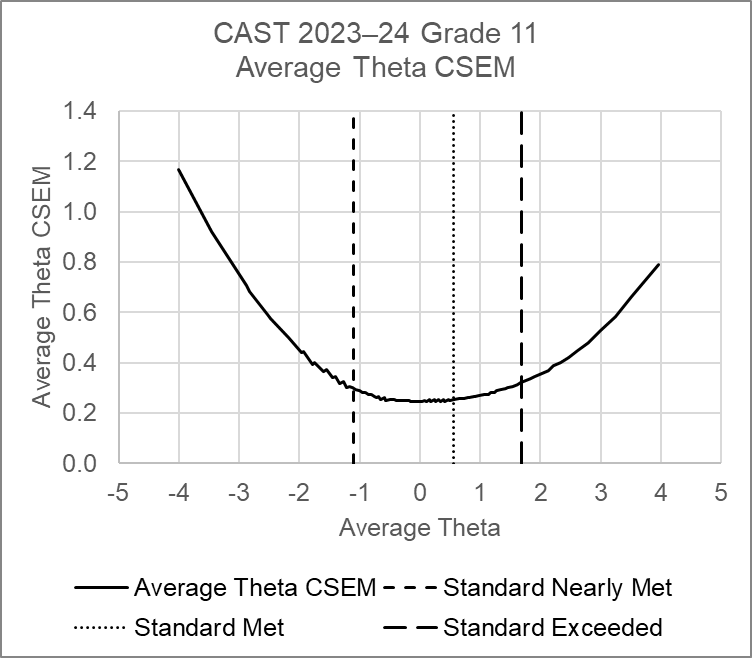


Figure 8.H.4 **Average CSEM by theta, grade eleven**

In table 8.H.4, rows where asterisks (\*) precede the values indicate the cut scores. These rows are also shaded.

Table 8.H.4 Average CSEM by Theta Data, Grade Eleven

|  |  |  |  |
| --- | --- | --- | --- |
| **Scale Score** | **N** | **Average Theta** | **Average Theta CSEM** |
| 553 | 2 | −4.00 | 1.17 |
| 554 | 17 | −3.45 | 0.92 |
| 556 | 78 | −2.88 | 0.71 |
| 557 | 26 | −2.83 | 0.68 |
| 558 | 30 | −2.52 | 0.59 |
| 559 | 316 | −2.47 | 0.58 |
| 561 | 793 | −2.18 | 0.50 |
| 563 | 778 | −1.97 | 0.44 |
| 564 | 953 | −1.93 | 0.44 |
| 565 | 384 | −1.79 | 0.39 |
| 566 | 2,873 | −1.74 | 0.40 |
| 568 | 3,066 | −1.60 | 0.36 |
| 569 | 2,076 | −1.55 | 0.37 |
| 570 | 3,316 | −1.46 | 0.34 |
| 571 | 3,866 | −1.40 | 0.34 |
| 572 | 3,172 | −1.33 | 0.32 |
| 573 | 5,890 | −1.28 | 0.32 |
| 574 | 3,617 | −1.21 | 0.30 |
| 575 | 3,885 | −1.17 | 0.30 |
| \*576 | \*5,490 | \*−1.12 | \*0.30 |
| 577 | 5,670 | −1.06 | 0.29 |
| 578 | 5,752 | −1.01 | 0.29 |
| 579 | 5,870 | −0.96 | 0.28 |
| 580 | 5,907 | −0.90 | 0.28 |
| 581 | 5,999 | −0.86 | 0.27 |
| 582 | 5,879 | −0.81 | 0.27 |
| 583 | 5,748 | −0.76 | 0.27 |
| 584 | 4,149 | −0.72 | 0.26 |
| 585 | 5,786 | −0.68 | 0.27 |
| 586 | 3,728 | −0.65 | 0.25 |
| 587 | 8,334 | −0.59 | 0.26 |
| 588 | 2,542 | −0.56 | 0.25 |
| 589 | 9,084 | −0.50 | 0.26 |
| 590 | 1,287 | −0.48 | 0.25 |
| 591 | 8,750 | −0.42 | 0.25 |
| 592 | 1,198 | −0.40 | 0.25 |
| 593 | 8,353 | −0.34 | 0.25 |
| 594 | 1,177 | −0.31 | 0.25 |
| 595 | 8,057 | −0.26 | 0.25 |
| 596 | 1,104 | −0.23 | 0.25 |
| 597 | 7,921 | −0.18 | 0.25 |
| 598 | 1,025 | −0.15 | 0.25 |
| 599 | 7,541 | −0.10 | 0.25 |
| 600 | 3,308 | −0.05 | 0.25 |
| 601 | 5,362 | −0.01 | 0.25 |
| 602 | 4,211 | 0.03 | 0.24 |
| 603 | 4,204 | 0.07 | 0.25 |
| 604 | 5,055 | 0.11 | 0.24 |
| 605 | 3,138 | 0.16 | 0.25 |
| 606 | 4,930 | 0.18 | 0.24 |
| 607 | 3,206 | 0.24 | 0.25 |
| 608 | 4,959 | 0.26 | 0.24 |
| 609 | 3,122 | 0.32 | 0.25 |
| 610 | 4,937 | 0.34 | 0.25 |
| 611 | 3,950 | 0.40 | 0.25 |
| 612 | 3,868 | 0.41 | 0.25 |
| 613 | 4,800 | 0.48 | 0.25 |
| 614 | 2,811 | 0.49 | 0.25 |
| \*615 | \*7,462 | \*0.56 | \*0.25 |
| 617 | 7,354 | 0.64 | 0.26 |
| 619 | 7,278 | 0.72 | 0.26 |
| 620 | 1,849 | 0.79 | 0.26 |
| 621 | 5,387 | 0.81 | 0.26 |
| 622 | 5,233 | 0.88 | 0.26 |
| 623 | 1,765 | 0.90 | 0.27 |
| 624 | 6,824 | 0.97 | 0.27 |
| 626 | 6,456 | 1.06 | 0.27 |
| 627 | 3,036 | 1.14 | 0.27 |
| 628 | 3,255 | 1.16 | 0.28 |
| 629 | 4,518 | 1.24 | 0.28 |
| 630 | 2,229 | 1.28 | 0.29 |
| 631 | 5,084 | 1.34 | 0.29 |
| 632 | 2,634 | 1.42 | 0.30 |
| 633 | 2,647 | 1.46 | 0.30 |
| 634 | 4,284 | 1.54 | 0.31 |
| 635 | 2,377 | 1.61 | 0.31 |
| \*636 | \*2,874 | \*1.67 | \*0.32 |
| 637 | 3,192 | 1.76 | 0.33 |
| 638 | 2,015 | 1.83 | 0.34 |
| 639 | 2,899 | 1.91 | 0.35 |
| 640 | 2,991 | 2.02 | 0.36 |
| 641 | 1,902 | 2.13 | 0.37 |
| 642 | 1,746 | 2.21 | 0.39 |
| 643 | 2,065 | 2.34 | 0.40 |
| 644 | 1,068 | 2.47 | 0.42 |
| 645 | 1,826 | 2.61 | 0.45 |
| 646 | 1,119 | 2.80 | 0.48 |
| 647 | 1,194 | 3.00 | 0.53 |
| 648 | 550 | 3.25 | 0.58 |
| 649 | 613 | 3.52 | 0.66 |
| 650 | 634 | 3.95 | 0.79 |

#### CSEMs for Grade Twelve

Figure 8.H.5 plots the average CSEMs by average theta score for grade twelve. The data used to create this graph is found in the table that immediately follows, table 8.H.5. The graph’s y-axis shows the average theta CSEMs corresponding to the theta values ranging from 0.0 to 1.4 in intervals of 0.2. The x-axis shows the average theta scores from −5 to +5 in intervals of 1.

Vertical lines in the CSEM curve represent the thresholds of *Standard Nearly Met*, *Standard Met*, and *Standard Exceeded* achievement levels. In general, for a given theta score, low values of the CSEM indicate greater measurement precision than that provided by larger CSEMs. The average theta scores to reach the *Standard Nearly Met*, *Standard Met*, and *Standard Exceeded* achievement levels are –1.12, 0.56, and 1.67, with average CSEMs of 0.30, 0.25, 0.32, respectively. As expected, the CSEM is highest at the extreme ends of the theta score distribution.

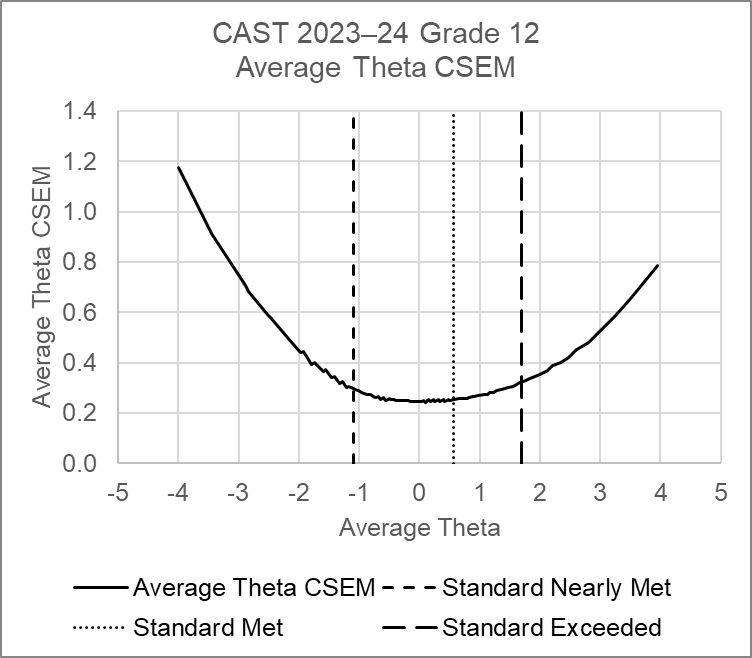


Figure 8.H.5 Average CSEM by theta, grade twelve

In table 8.H.5, rows where asterisks (\*) precede the values indicate the cut scores. These rows are also shaded.

Table 8.H.5 Average CSEM by Theta Data, Grade Twelve

|  |  |  |  |
| --- | --- | --- | --- |
| **Scale Score** | **N** | **Average Theta** | **Average Theta CSEM** |
| 553 | 4 | −4.00 | 1.17 |
| 554 | 8 | −3.45 | 0.91 |
| 556 | 16 | −2.88 | 0.71 |
| 557 | 15 | −2.84 | 0.68 |
| 558 | 16 | −2.52 | 0.59 |
| 559 | 142 | −2.47 | 0.58 |
| 561 | 359 | −2.18 | 0.50 |
| 563 | 328 | −1.97 | 0.44 |
| 564 | 438 | −1.93 | 0.44 |
| 565 | 139 | −1.79 | 0.39 |
| 566 | 1,233 | −1.74 | 0.40 |
| 568 | 1,259 | −1.60 | 0.36 |
| 569 | 875 | −1.55 | 0.37 |
| 570 | 1,348 | −1.46 | 0.34 |
| 571 | 1,540 | −1.40 | 0.34 |
| 572 | 1,203 | −1.33 | 0.32 |
| 573 | 2,103 | −1.28 | 0.32 |
| 574 | 1,386 | −1.21 | 0.30 |
| 575 | 1,416 | −1.17 | 0.31 |
| \*576 | \*2,000 | \*−1.12 | \*0.30 |
| 577 | 2,061 | −1.06 | 0.29 |
| 578 | 2,153 | −1.01 | 0.29 |
| 579 | 2,175 | −0.96 | 0.28 |
| 580 | 2,067 | −0.91 | 0.28 |
| 581 | 2,038 | −0.86 | 0.27 |
| 582 | 2,045 | −0.81 | 0.27 |
| 583 | 1,963 | −0.76 | 0.27 |
| 584 | 1,382 | −0.72 | 0.26 |
| 585 | 1,929 | −0.68 | 0.27 |
| 586 | 1,250 | −0.65 | 0.25 |
| 587 | 2,719 | −0.59 | 0.26 |
| 588 | 829 | −0.56 | 0.25 |
| 589 | 2,875 | −0.50 | 0.26 |
| 590 | 393 | −0.48 | 0.25 |
| 591 | 2,713 | −0.42 | 0.25 |
| 592 | 374 | −0.40 | 0.25 |
| 593 | 2,573 | −0.34 | 0.25 |
| 594 | 342 | −0.31 | 0.25 |
| 595 | 2,519 | −0.25 | 0.25 |
| 596 | 353 | −0.23 | 0.25 |
| 597 | 2,337 | −0.18 | 0.25 |
| 598 | 319 | −0.15 | 0.25 |
| 599 | 2,203 | −0.10 | 0.25 |
| 600 | 942 | −0.05 | 0.25 |
| 601 | 1,515 | −0.01 | 0.25 |
| 602 | 1,193 | 0.03 | 0.24 |
| 603 | 1,148 | 0.07 | 0.25 |
| 604 | 1,385 | 0.11 | 0.24 |
| 605 | 903 | 0.16 | 0.25 |
| 606 | 1,408 | 0.18 | 0.24 |
| 607 | 883 | 0.24 | 0.25 |
| 608 | 1,341 | 0.26 | 0.24 |
| 609 | 788 | 0.32 | 0.25 |
| 610 | 1,228 | 0.33 | 0.25 |
| 611 | 1,027 | 0.40 | 0.25 |
| 612 | 937 | 0.41 | 0.25 |
| 613 | 1,244 | 0.48 | 0.25 |
| 614 | 701 | 0.49 | 0.25 |
| \*615 | \*1,916 | \*0.56 | \*0.25 |
| 617 | 1,897 | 0.64 | 0.26 |
| 619 | 1,766 | 0.72 | 0.26 |
| 620 | 446 | 0.79 | 0.26 |
| 621 | 1,316 | 0.81 | 0.26 |
| 622 | 1,268 | 0.88 | 0.27 |
| 623 | 411 | 0.90 | 0.27 |
| 624 | 1,659 | 0.97 | 0.27 |
| 626 | 1,587 | 1.06 | 0.27 |
| 627 | 744 | 1.14 | 0.27 |
| 628 | 771 | 1.16 | 0.28 |
| 629 | 1,148 | 1.24 | 0.28 |
| 630 | 529 | 1.28 | 0.29 |
| 631 | 1,283 | 1.34 | 0.29 |
| 632 | 623 | 1.42 | 0.30 |
| 633 | 625 | 1.46 | 0.30 |
| 634 | 1,020 | 1.54 | 0.31 |
| 635 | 563 | 1.61 | 0.32 |
| \*636 | \*704 | \*1.67 | \*0.32 |
| 637 | 775 | 1.76 | 0.33 |
| 638 | 509 | 1.83 | 0.34 |
| 639 | 704 | 1.91 | 0.35 |
| 640 | 775 | 2.02 | 0.36 |
| 641 | 495 | 2.13 | 0.37 |
| 642 | 477 | 2.21 | 0.39 |
| 643 | 594 | 2.34 | 0.40 |
| 644 | 262 | 2.47 | 0.42 |
| 645 | 532 | 2.61 | 0.45 |
| 646 | 287 | 2.80 | 0.48 |
| 647 | 350 | 3.00 | 0.53 |
| 648 | 158 | 3.24 | 0.58 |
| 649 | 183 | 3.51 | 0.66 |
| 650 | 189 | 3.95 | 0.79 |

#### CSEMs for the High School Grade Band

Figure 8.H.6 plots the average CSEMs by average theta score for high school. The data used to create this graph is found in the table that immediately follows, table 8.H.6. The graph’s y-axis shows the average theta CSEMs corresponding to the theta values ranging from 0.0 to 1.4 in intervals of 0.2. The x-axis shows the average theta scores from −5 to +5 in intervals of 1.

Vertical lines in the CSEM curve represent the thresholds of *Standard Nearly Met*, *Standard Met*, and *Standard Exceeded* achievement levels. In general, for a given theta score, low values of the CSEM indicate greater measurement precision than that provided by larger CSEMs. The average theta scores to reach the *Standard Nearly Met*, *Standard Met*, and *Standard Exceeded* achievement levels are –1.12, 0.56, and 1.67, with average CSEMs of 0.30, 0.25, 0.32, respectively. As expected, the CSEM is highest at the extreme ends of the theta score distribution.

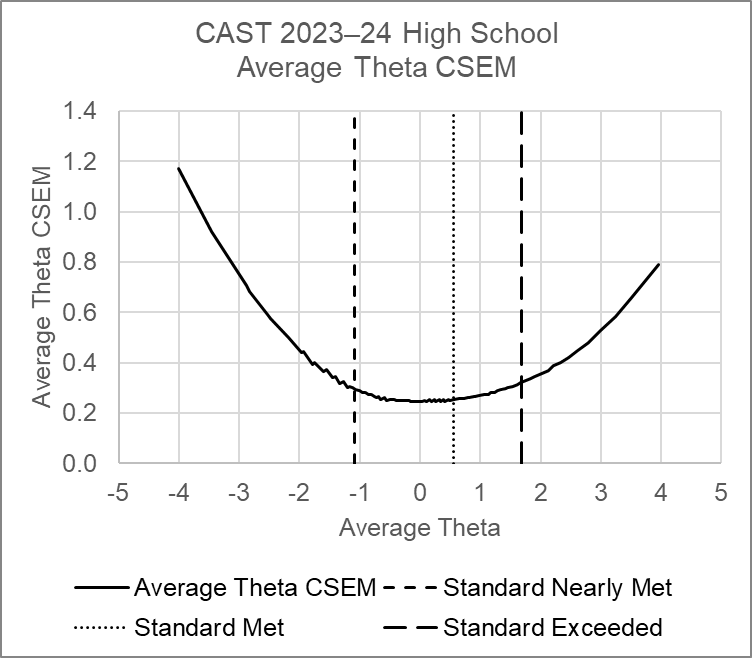


Figure 8.H.6 Average CSEM by theta, high school

In table 8.H.6, rows where asterisks (\*) precede the values indicate the cut scores. These rows are also shaded.

Table 8.H.6 Average CSEM by Theta Data, High School

|  |  |  |  |
| --- | --- | --- | --- |
| **Scale Score** | **N** | **Average Theta** | **Average Theta CSEM** |
| 553 | 7 | −4.00 | 1.17 |
| 554 | 27 | −3.45 | 0.92 |
| 556 | 103 | −2.88 | 0.71 |
| 557 | 42 | −2.83 | 0.68 |
| 558 | 50 | −2.52 | 0.59 |
| 559 | 480 | −2.47 | 0.58 |
| 561 | 1,206 | −2.18 | 0.50 |
| 563 | 1,166 | −1.97 | 0.44 |
| 564 | 1,483 | −1.93 | 0.44 |
| 565 | 545 | −1.79 | 0.39 |
| 566 | 4,326 | −1.74 | 0.40 |
| 568 | 4,530 | −1.60 | 0.36 |
| 569 | 3,097 | −1.55 | 0.37 |
| 570 | 4,939 | −1.46 | 0.34 |
| 571 | 5,723 | −1.40 | 0.34 |
| 572 | 4,674 | −1.33 | 0.32 |
| 573 | 8,438 | −1.28 | 0.32 |
| 574 | 5,327 | −1.21 | 0.30 |
| 575 | 5,612 | −1.17 | 0.31 |
| \*576 | \*7,970 | \*−1.12 | \*0.30 |
| 577 | 8,194 | −1.06 | 0.29 |
| 578 | 8,424 | −1.01 | 0.29 |
| 579 | 8,551 | −0.96 | 0.28 |
| 580 | 8,490 | −0.90 | 0.28 |
| 581 | 8,546 | −0.86 | 0.27 |
| 582 | 8,445 | −0.81 | 0.27 |
| 583 | 8,240 | −0.76 | 0.27 |
| 584 | 5,906 | −0.72 | 0.26 |
| 585 | 8,205 | −0.68 | 0.27 |
| 586 | 5,335 | −0.65 | 0.25 |
| 587 | 11,786 | −0.59 | 0.26 |
| 588 | 3,587 | −0.56 | 0.25 |
| 589 | 12,776 | −0.50 | 0.26 |
| 590 | 1,815 | −0.48 | 0.25 |
| 591 | 12,219 | −0.42 | 0.25 |
| 592 | 1,679 | −0.40 | 0.25 |
| 593 | 11,617 | −0.34 | 0.25 |
| 594 | 1,614 | −0.31 | 0.25 |
| 595 | 11,275 | −0.26 | 0.25 |
| 596 | 1,547 | −0.23 | 0.25 |
| 597 | 10,919 | −0.18 | 0.25 |
| 598 | 1,442 | −0.15 | 0.25 |
| 599 | 10,363 | −0.10 | 0.25 |
| 600 | 4,529 | −0.05 | 0.25 |
| 601 | 7,265 | −0.01 | 0.25 |
| 602 | 5,761 | 0.03 | 0.24 |
| 603 | 5,716 | 0.07 | 0.25 |
| 604 | 6,824 | 0.11 | 0.24 |
| 605 | 4,284 | 0.16 | 0.25 |
| 606 | 6,718 | 0.18 | 0.24 |
| 607 | 4,324 | 0.24 | 0.25 |
| 608 | 6,670 | 0.26 | 0.24 |
| 609 | 4,157 | 0.32 | 0.25 |
| 610 | 6,521 | 0.34 | 0.25 |
| 611 | 5,285 | 0.40 | 0.25 |
| 612 | 5,058 | 0.41 | 0.25 |
| 613 | 6,416 | 0.48 | 0.25 |
| 614 | 3,708 | 0.49 | 0.25 |
| \*615 | \*9,883 | \*0.56 | \*0.25 |
| 617 | 9,780 | 0.64 | 0.26 |
| 619 | 9,531 | 0.72 | 0.26 |
| 620 | 2,403 | 0.79 | 0.26 |
| 621 | 7,041 | 0.81 | 0.26 |
| 622 | 6,825 | 0.88 | 0.27 |
| 623 | 2,281 | 0.90 | 0.27 |
| 624 | 8,915 | 0.97 | 0.27 |
| 626 | 8,458 | 1.06 | 0.27 |
| 627 | 3,954 | 1.14 | 0.27 |
| 628 | 4,227 | 1.16 | 0.28 |
| 629 | 5,895 | 1.24 | 0.28 |
| 630 | 2,885 | 1.28 | 0.29 |
| 631 | 6,638 | 1.34 | 0.29 |
| 632 | 3,407 | 1.42 | 0.30 |
| 633 | 3,424 | 1.46 | 0.30 |
| 634 | 5,524 | 1.54 | 0.31 |
| 635 | 3,073 | 1.61 | 0.31 |
| \*636 | \*3,730 | \*1.67 | \*0.32 |
| 637 | 4,125 | 1.76 | 0.33 |
| 638 | 2,621 | 1.83 | 0.34 |
| 639 | 3,736 | 1.91 | 0.35 |
| 640 | 3,883 | 2.02 | 0.36 |
| 641 | 2,475 | 2.13 | 0.37 |
| 642 | 2,296 | 2.21 | 0.39 |
| 643 | 2,774 | 2.34 | 0.40 |
| 644 | 1,384 | 2.47 | 0.42 |
| 645 | 2,440 | 2.61 | 0.45 |
| 646 | 1,444 | 2.80 | 0.48 |
| 647 | 1,600 | 3.00 | 0.53 |
| 648 | 729 | 3.24 | 0.58 |
| 649 | 822 | 3.52 | 0.66 |
| 650 | 851 | 3.95 | 0.79 |

### Appendix 8.I: Analyses of Classification

**Notes:**

* For the tables on decision accuracy, the horizontal headers indicate values of classification by all-forms average, while the vertical headers indicate values of classification by form taken category. For example, the value 0.14 in the *150–178* row in the *Standard Not Met* column in table 8.I.1 means that 14 percent of students were classified as Standard Not Met by all-forms average and classified as Standard Not Met (placement score within a range of 150 and 178) by their observed scores.
* The proportion of students accurately classified across all achievement levels (that is, the first number reported under each table on decision accuracy) is the sum of the main diagonal elements of the decision accuracy tables. Using the threshold of Standard Met, the classifications are collapsed to *Standard Not Met* and *Standard Nearly Met* versus *Standard Met* and *Standard Exceeded*. The proportion of students accurately classified in the two collapsed categories (that is, the second number reported under each table on decision accuracy) is the sum of the main diagonal elements of the collapsed two-by-two decision accuracy tables.
* For the tables on decision consistency, the horizontal headers indicate values of classification by alternate form while the vertical headers indicate values of classification by form taken category. For example, the value 0.13 in the *150–178* row in the *Standard Not Met* column in table 8.I.2 means that 13 percent of students were classified as Standard Not Met by alternate form and classified as Standard Not Met (placement score within a range of 150 and 178) by their observed scores.
* The proportion of students consistently classified across all achievement levels (that is, the first number reported under each table on decision consistency) is the sum of the main diagonal elements of the decision consistency tables. Using the threshold of Standard Met, the classifications are collapsed to *Standard Not Met* and *Standard Nearly Met* versus *Standard Met* and *Standard Exceeded*. The proportion of students consistently classified in the two collapsed categories (that is, the second number reported under each table on decision consistency) is the sum of the main diagonal elements of the collapsed two-by-two decision consistency tables.
* The sum of the main diagonal elements may not exactly match the values under the tables because of rounding.

Table 8.I.1 Grade Five: Decision Accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Placement Score** | **Standard Not Met** | **Standard Nearly Met** | **Standard Met** | **Standard Exceeded** | **Category Total** |
| 150–178 | 0.14 | 0.04 | 0.00 | 0.00 | 0.18 |
| 179–213 | 0.03 | 0.43 | 0.03 | 0.00 | 0.49 |
| 214–230 | 0.00 | 0.04 | 0.14 | 0.02 | 0.21 |
| 231–250 | 0.00 | 0.00 | 0.03 | 0.10 | 0.12 |

All-forms average, estimated proportion accurately classified: four achievement levels = 0.80; two collapsed categories = 0.93

Table 8.I.2 Grade Five: Decision Consistency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Placement Score** | **Standard Not Met** | **Standard Nearly Met** | **Standard Met** | **Standard Exceeded** | **Category Total** |
| 150–178 | 0.13 | 0.05 | 0.00 | 0.00 | 0.18 |
| 179–213 | 0.05 | 0.39 | 0.05 | 0.00 | 0.49 |
| 214–230 | 0.00 | 0.05 | 0.12 | 0.03 | 0.21 |
| 231–250 | 0.00 | 0.00 | 0.03 | 0.09 | 0.12 |

Alternate form, estimated proportion consistently classified: four achievement levels = 0.73; two collapsed categories = 0.90

Table 8.I.3 Grade Eight: Decision Accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Placement Score** | **Standard Not Met** | **Standard Nearly Met** | **Standard Met** | **Standard Exceeded** | **Category Total** |
| 350–377 | 0.12 | 0.04 | 0.00 | 0.00 | 0.16 |
| 378–414 | 0.04 | 0.48 | 0.03 | 0.00 | 0.55 |
| 415–432 | 0.00 | 0.04 | 0.14 | 0.01 | 0.19 |
| 433–450 | 0.00 | 0.00 | 0.02 | 0.08 | 0.10 |

All-forms average, estimated proportion accurately classified: four achievement levels = 0.81; two collapsed categories = 0.93

Table 8.I.4 Grade Eight: Decision Consistency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Placement Score** | **Standard Not Met** | **Standard Nearly Met** | **Standard Met** | **Standard Exceeded** | **Category Total** |
| 350–377 | 0.11 | 0.05 | 0.00 | 0.00 | 0.16 |
| 378–414 | 0.06 | 0.44 | 0.05 | 0.00 | 0.55 |
| 415–432 | 0.00 | 0.05 | 0.12 | 0.03 | 0.19 |
| 433–450 | 0.00 | 0.00 | 0.02 | 0.07 | 0.10 |

Alternate form, estimated proportion consistently classified: four achievement levels = 0.74; two collapsed categories = 0.90

Table 8.I.5 High School—Grade Ten: Decision Accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Placement Score** | **Standard Not Met** | **Standard Nearly Met** | **Standard Met** | **Standard Exceeded** | **Category Total** |
| 550–575 | 0.07 | 0.05 | 0.00 | 0.00 | 0.12 |
| 576–614 | 0.04 | 0.57 | 0.03 | 0.00 | 0.64 |
| 615–635 | 0.00 | 0.04 | 0.15 | 0.01 | 0.20 |
| 636–650 | 0.00 | 0.00 | 0.01 | 0.04 | 0.05 |

All-forms average, estimated proportion accurately classified: four achievement levels = 0.83; two collapsed categories = 0.94

Table 8.I.6 High School—Grade Ten: Decision Consistency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Placement Score** | **Standard Not Met** | **Standard Nearly Met** | **Standard Met** | **Standard Exceeded** | **Category Total** |
| 550–575 | 0.06 | 0.05 | 0.00 | 0.00 | 0.12 |
| 576–614 | 0.06 | 0.53 | 0.04 | 0.00 | 0.64 |
| 615–635 | 0.00 | 0.05 | 0.13 | 0.02 | 0.20 |
| 636–650 | 0.00 | 0.00 | 0.02 | 0.03 | 0.05 |

Alternate form, estimated proportion consistently classified: four achievement levels = 0.76; two collapsed categories = 0.91

Table 8.I.7 High School—Grade Eleven: Decision Accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Placement Score** | **Standard Not Met** | **Standard Nearly Met** | **Standard Met** | **Standard Exceeded** | **Category Total** |
| 550–575 | 0.07 | 0.04 | 0.00 | 0.00 | 0.11 |
| 576–614 | 0.03 | 0.51 | 0.03 | 0.00 | 0.57 |
| 615–635 | 0.00 | 0.04 | 0.18 | 0.02 | 0.24 |
| 636–650 | 0.00 | 0.00 | 0.02 | 0.06 | 0.08 |

All-forms average, estimated proportion accurately classified: four achievement levels = 0.82; two collapsed categories = 0.93

Table 8.I.8 High School—Grade Eleven: Decision Consistency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Placement Score** | **Standard Not Met** | **Standard Nearly Met** | **Standard Met** | **Standard Exceeded** | **Category Total** |
| 550–575 | 0.06 | 0.05 | 0.00 | 0.00 | 0.11 |
| 576–614 | 0.05 | 0.47 | 0.04 | 0.00 | 0.57 |
| 615–635 | 0.00 | 0.05 | 0.16 | 0.03 | 0.24 |
| 636–650 | 0.00 | 0.00 | 0.02 | 0.06 | 0.08 |

Alternate form, estimated proportion consistently classified: four achievement levels = 0.75; two collapsed categories = 0.90

Table 8.I.9 High School—Grade Twelve: Decision Accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Placement Score** | **Standard Not Met** | **Standard Nearly Met** | **Standard Met** | **Standard Exceeded** | **Category Total** |
| 550–575 | 0.10 | 0.05 | 0.00 | 0.00 | 0.14 |
| 576–614 | 0.05 | 0.52 | 0.02 | 0.00 | 0.59 |
| 615–635 | 0.00 | 0.03 | 0.15 | 0.01 | 0.20 |
| 636–650 | 0.00 | 0.00 | 0.02 | 0.05 | 0.07 |

All-forms average, estimated proportion accurately classified: four achievement levels = 0.81; two collapsed categories = 0.94

Table 8.I.10 High School—Grade Twelve: Decision Consistency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Placement Score** | **Standard Not Met** | **Standard Nearly Met** | **Standard Met** | **Standard Exceeded** | **Category Total** |
| 550–575 | 0.08 | 0.06 | 0.00 | 0.00 | 0.14 |
| 576–614 | 0.07 | 0.48 | 0.04 | 0.00 | 0.59 |
| 615–635 | 0.00 | 0.04 | 0.13 | 0.02 | 0.20 |
| 636–650 | 0.00 | 0.00 | 0.02 | 0.05 | 0.07 |

Alternate form, estimated proportion consistently classified: four achievement levels = 0.74; two collapsed categories = 0.92

Table 8.I.11 High School: Decision Accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Placement Score** | **Standard Not Met** | **Standard Nearly Met** | **Standard Met** | **Standard Exceeded** | **Category Total** |
| 550–575 | 0.08 | 0.04 | 0.00 | 0.00 | 0.12 |
| 576–614 | 0.04 | 0.51 | 0.03 | 0.00 | 0.58 |
| 615–635 | 0.00 | 0.04 | 0.17 | 0.01 | 0.23 |
| 636–650 | 0.00 | 0.00 | 0.02 | 0.06 | 0.08 |

All-forms average, estimated proportion accurately classified: four achievement levels = 0.82; two collapsed categories = 0.93

Table 8.I.12 High School: Decision Consistency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Placement Score** | **Standard Not Met** | **Standard Nearly Met** | **Standard Met** | **Standard Exceeded** | **Category Total** |
| 550–575 | 0.07 | 0.05 | 0.00 | 0.00 | 0.12 |
| 576–614 | 0.06 | 0.48 | 0.04 | 0.00 | 0.58 |
| 615–635 | 0.00 | 0.05 | 0.15 | 0.02 | 0.23 |
| 636–650 | 0.00 | 0.00 | 0.02 | 0.05 | 0.08 |

Alternate form, estimated proportion consistently classified: four achievement levels = 0.75; two collapsed categories = 0.91

### Appendix 8.J: Interrater Reliability

Table 8.J.1 Interrater Reliability and Descriptive Statistics for the Ratings by AI and Human Raters in AI Scoring of Operational CR Items for Grade Five

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Prompt** | **Item ID** | **Score Points** | **Number of Responses** | **Kappa** | **QWK** | **Percent Exact** | **Percent Adjacent** | **Percent Exact + Adjacent** | **AI Rater Item Score Mean** | **AI Rater Item Score SD** | **Human Rater Item Score Mean** | **Human Rater Item Score SD** | **SMD** |
| 1 | VH831833 | 2 | 1,771 | 0.75 | 0.84 | 83.40 | 15.70 | 99.10 | 1.06 | 0.77 | 1.01 | 0.80 | 0.07 |
| 2 | VH837166 | 2 | 1,765 | 0.71 | 0.82 | 81.53 | 18.19 | 99.72 | 0.93 | 0.71 | 0.99 | 0.76 | −0.07 |
| 3 | VH882648 | 2 | 1,778 | 0.73 | 0.87 | 87.18 | 12.71 | 99.89 | 0.41 | 0.68 | 0.47 | 0.73 | −0.07 |
| 4 | VH882701 | 2 | 1,779 | 0.56 | 0.77 | 73.13 | 26.42 | 99.55 | 0.66 | 0.74 | 0.70 | 0.82 | −0.05 |
| 5 | VR138786 | 2 | 1,779 | 0.67 | 0.79 | 85.33 | 14.33 | 99.66 | 0.34 | 0.57 | 0.37 | 0.64 | −0.05 |
| 6 | VR159560 | 2 | 1,772 | 0.56 | 0.71 | 71.50 | 27.77 | 99.27 | 1.10 | 0.69 | 1.03 | 0.77 | 0.09 |
| 7 | VR292950 | 2 | 1,770 | 0.64 | 0.84 | 77.12 | 22.32 | 99.44 | 0.84 | 0.83 | 0.85 | 0.91 | −0.01 |
| 8 | VR298143 | 2 | 1,766 | 0.61 | 0.78 | 77.86 | 21.74 | 99.60 | 0.57 | 0.70 | 0.57 | 0.76 | 0.00 |
| **N/A** | **AVERAGE:** | **N/A** | **1,773** | **0.65** | **0.80** | **79.63** | **19.90** | **99.53** | **0.74** | **0.71** | **0.75** | **0.77** | **−0.01** |

Table 8.J.2 Interrater Reliability and Descriptive Statistics for the Ratings by AI and Human Raters in AI Scoring of Operational CR Items for Grade Eight

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Prompt** | **Item ID** | **Score Points** | **Number of Responses** | **Kappa** | **QWK** | **Percent Exact** | **Percent Adjacent** | **Percent Exact + Adjacent** | **AI Rater Item Score Mean** | **AI Rater Item Score SD** | **Human Rater Item Score Mean** | **Human Rater Item Score SD** | **SMD** |
| 1 | VH834406 | 2 | 1,789 | 0.54 | 0.77 | 69.54 | 29.96 | 99.50 | 0.80 | 0.77 | 0.82 | 0.91 | −0.02 |
| 2 | VH837853 | 2 | 1,777 | 0.70 | 0.86 | 80.98 | 18.85 | 99.83 | 0.85 | 0.85 | 0.84 | 0.85 | 0.01 |
| 3 | VH858277 | 2 | 1,777 | 0.70 | 0.85 | 80.70 | 19.08 | 99.77 | 0.83 | 0.79 | 0.81 | 0.84 | 0.02 |
| 4 | VH882560 | 2 | 1,786 | 0.77 | 0.90 | 85.55 | 14.17 | 99.72 | 1.20 | 0.87 | 1.17 | 0.90 | 0.03 |
| 5 | VH882650 | 2 | 1,782 | 0.86 | 0.94 | 92.03 | 7.80 | 99.83 | 0.74 | 0.86 | 0.72 | 0.88 | 0.02 |
| 6 | VH882698 | 2 | 1,781 | 0.69 | 0.81 | 82.26 | 17.63 | 99.89 | 0.55 | 0.69 | 0.56 | 0.70 | −0.02 |
| 7 | VR024537 | 2 | 1,783 | 0.67 | 0.83 | 78.52 | 21.03 | 99.55 | 0.85 | 0.79 | 0.85 | 0.84 | −0.01 |
| 8 | VR026008 | 2 | 1,786 | 0.65 | 0.83 | 82.47 | 17.13 | 99.61 | 0.48 | 0.71 | 0.51 | 0.80 | −0.03 |
| 9 | VR301062 | 2 | 1,775 | 0.64 | 0.81 | 80.28 | 19.44 | 99.72 | 0.50 | 0.69 | 0.56 | 0.79 | −0.07 |
| **N/A** | **AVERAGE:** | **N/A** | **1,782** | **0.69** | **0.84** | **81.37** | **18.34** | **99.71** | **0.76** | **0.78** | **0.76** | **0.83** | **−0.01** |

Table 8.J.3 Interrater Reliability and Descriptive Statistics for the Ratings by AI and Human Raters in AI Scoring of Operational CR Items for High School

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Prompt** | **Item ID** | **Score Points** | **Number of Students** | **Kappa** | **QWK** | **Percent Exact** | **Percent Adjacent** | **Percent Exact + Adjacent** | **AI Rater Item Score Mean** | **AI Rater Item Score SD** | **Human Rater Item Score Mean** | **Human Rater Item Score SD** | **SMD** |
| 1 | VH849536 | 2 | 1,791 | 0.63 | 0.76 | 78.56 | 20.99 | 99.55 | 0.59 | 0.68 | 0.62 | 0.71 | −0.03 |
| 2 | VH852204 | 2 | 1,784 | 0.63 | 0.79 | 80.04 | 19.79 | 99.83 | 0.51 | 0.68 | 0.53 | 0.73 | −0.03 |
| 3 | VH870782 | 2 | 1,798 | 0.55 | 0.68 | 81.59 | 18.13 | 99.72 | 0.31 | 0.52 | 0.31 | 0.58 | −0.01 |
| 4 | VH876098 | 2 | 1,785 | 0.59 | 0.80 | 72.61 | 26.83 | 99.44 | 0.97 | 0.81 | 0.99 | 0.89 | −0.03 |
| 5 | VH876156 | 2 | 1,769 | 0.78 | 0.88 | 85.53 | 14.47 | 100.00 | 0.81 | 0.76 | 0.77 | 0.80 | 0.05 |
| 6 | VR009080 | 2 | 1,781 | 0.64 | 0.82 | 79.17 | 20.61 | 99.78 | 0.57 | 0.72 | 0.65 | 0.80 | −0.09 |
| 7 | VR117153 | 2 | 1,787 | 0.63 | 0.79 | 78.18 | 21.77 | 99.94 | 0.58 | 0.69 | 0.60 | 0.76 | −0.03 |
| 8 | VR152036 | 2 | 1,782 | 0.70 | 0.85 | 81.20 | 18.46 | 99.66 | 0.72 | 0.77 | 0.78 | 0.84 | −0.08 |
| 9 | VR306216 | 2 | 1,788 | 0.61 | 0.80 | 76.01 | 23.55 | 99.55 | 0.64 | 0.76 | 0.72 | 0.84 | −0.09 |
| **N/A** | **AVERAGE:** | **N/A** | **1,785** | **0.64** | **0.80** | **79.21** | **20.51** | **99.72** | **0.63** | **0.71** | **0.66** | **0.77** | **−0.04** |

### Appendix 8.K: Correlations to Smarter Balanced Test Scores

Table 8.K.1 Correlations with Smarter Balanced ELA Test Scores for Grade Five

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number Tested** | **Correlation** |
| All students | 417,138 | 0.84 |
| Male | 212,800 | 0.85 |
| Female | 204,223 | 0.84 |
| Nonbinary | 115 | 0.88 |
| EL | 76,059 | 0.69 |
| English only | 259,098 | 0.84 |
| RFEP | 55,458 | 0.77 |
| IFEP | 26,433 | 0.82 |
| ADEL | 0 | N/A |
| To be determined | 27 | 0.83 |
| English proficiency unknown | 63 | 0.84 |
| Economically disadvantaged | 273,965 | 0.81 |
| Not economically disadvantaged | 143,173 | 0.83 |
| American Indian or Alaska Native (All) | 1,792 | 0.80 |
| Asian (All) | 44,096 | 0.85 |
| Native Hawaiian or Other Pacific Islander (All) | 1,696 | 0.82 |
| Filipino (All) | 9,358 | 0.81 |
| Hispanic or Latino (All) | 231,620 | 0.81 |
| Black or African American (All) | 20,760 | 0.81 |
| White (All) | 83,740 | 0.82 |
| Two or more races (All) | 24,076 | 0.84 |
| Disability | 56,331 | 0.79 |
| No disability | 360,807 | 0.83 |
| Migrant education | 3,291 | 0.79 |
| Not migrant education | 413,847 | 0.84 |
| Armed forces family member | 6,470 | 0.83 |
| Not armed forces family member | 410,668 | 0.84 |
| Homeless | 14,740 | 0.80 |
| Not homeless | 402,398 | 0.84 |
| Foster youth | 1,344 | 0.77 |
| Not foster youth | 415,794 | 0.84 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 464 | 0.82 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 1,328 | 0.77 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 25,210 | 0.81 |
| Asian (Primary ethnicity—Economically disadvantaged) | 18,886 | 0.85 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 422 | 0.81 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 1,274 | 0.81 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 5,292 | 0.80 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 4,066 | 0.81 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 44,398 | 0.82 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 187,222 | 0.80 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 4,112 | 0.82 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 16,648 | 0.79 |
| White (Primary ethnicity—Not economically disadvantaged) | 50,065 | 0.80 |
| White (Primary ethnicity—Economically disadvantaged) | 33,675 | 0.82 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 13,210 | 0.81 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 10,866 | 0.83 |

Table 8.K.2 Correlations with Smarter Balanced Mathematics Test Scores for Grade Five

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number Tested** | **Correlation** |
| All students | 421,290 | 0.81 |
| Male | 214,976 | 0.82 |
| Female | 206,198 | 0.81 |
| Nonbinary | 116 | 0.83 |
| EL | 80,380 | 0.63 |
| English only | 258,914 | 0.81 |
| RFEP | 55,426 | 0.76 |
| IFEP | 26,438 | 0.81 |
| ADEL | 0 | N/A |
| To be determined | 69 | 0.77 |
| English proficiency unknown | 63 | 0.85 |
| Economically disadvantaged | 276,467 | 0.77 |
| Not economically disadvantaged | 144,823 | 0.81 |
| American Indian or Alaska Native (All) | 1,789 | 0.78 |
| Asian (All) | 44,880 | 0.82 |
| Native Hawaiian or Other Pacific Islander (All) | 1,698 | 0.77 |
| Filipino (All) | 9,402 | 0.79 |
| Hispanic or Latino (All) | 234,349 | 0.77 |
| Black or African American (All) | 20,777 | 0.76 |
| White (All) | 84,210 | 0.80 |
| Two or more races (All) | 24,185 | 0.82 |
| Disability | 56,264 | 0.76 |
| No disability | 365,026 | 0.80 |
| Migrant education | 3,424 | 0.73 |
| Not migrant education | 417,866 | 0.81 |
| Armed forces family member | 6,482 | 0.80 |
| Not armed forces family member | 414,808 | 0.81 |
| Homeless | 15,478 | 0.75 |
| Not homeless | 405,812 | 0.81 |
| Foster youth | 1,344 | 0.70 |
| Not foster youth | 419,946 | 0.81 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 464 | 0.78 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 1,325 | 0.75 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 25,650 | 0.78 |
| Asian (Primary ethnicity—Economically disadvantaged) | 19,230 | 0.82 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 422 | 0.77 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 1,276 | 0.76 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 5,311 | 0.78 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 4,091 | 0.79 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 45,366 | 0.80 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 188,983 | 0.75 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 4,119 | 0.79 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 16,658 | 0.73 |
| White (Primary ethnicity—Not economically disadvantaged) | 50,207 | 0.77 |
| White (Primary ethnicity—Economically disadvantaged) | 34,003 | 0.78 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 13,284 | 0.80 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 10,901 | 0.80 |

Table 8.K.3 Correlations with Smarter Balanced ELA Test Scores for Grade Eight

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number Tested** | **Correlation** |
| All students | 421,734 | 0.81 |
| Male | 216,025 | 0.82 |
| Female | 205,334 | 0.81 |
| Nonbinary | 375 | 0.77 |
| EL | 50,853 | 0.54 |
| English only | 241,522 | 0.81 |
| RFEP | 111,162 | 0.76 |
| IFEP | 18,133 | 0.80 |
| ADEL | 0 | N/A |
| To be determined | 32 | 0.86 |
| English proficiency unknown | 32 | 0.73 |
| Economically disadvantaged | 274,383 | 0.78 |
| Not economically disadvantaged | 147,351 | 0.81 |
| American Indian or Alaska Native (All) | 1,784 | 0.78 |
| Asian (All) | 41,880 | 0.82 |
| Native Hawaiian or Other Pacific Islander (All) | 1,826 | 0.78 |
| Filipino (All) | 10,021 | 0.79 |
| Hispanic or Latino (All) | 238,850 | 0.77 |
| Black or African American (All) | 21,191 | 0.76 |
| White (All) | 84,544 | 0.80 |
| Two or more races (All) | 21,638 | 0.82 |
| Disability | 50,036 | 0.73 |
| No disability | 371,698 | 0.80 |
| Migrant education | 3,242 | 0.76 |
| Not migrant education | 418,492 | 0.81 |
| Armed forces family member | 5,943 | 0.80 |
| Not armed forces family member | 415,791 | 0.81 |
| Homeless | 13,076 | 0.76 |
| Not homeless | 408,658 | 0.81 |
| Foster youth | 1,400 | 0.71 |
| Not foster youth | 420,334 | 0.81 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 520 | 0.80 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 1,264 | 0.75 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 24,317 | 0.78 |
| Asian (Primary ethnicity—Economically disadvantaged) | 17,563 | 0.82 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 511 | 0.82 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 1,315 | 0.74 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 5,902 | 0.78 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 4,119 | 0.78 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 47,522 | 0.79 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 191,328 | 0.76 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 4,855 | 0.78 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 16,336 | 0.74 |
| White (Primary ethnicity—Not economically disadvantaged) | 51,407 | 0.78 |
| White (Primary ethnicity—Economically disadvantaged) | 33,137 | 0.79 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 12,317 | 0.79 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 9,321 | 0.79 |

Table 8.K.4 Correlations with Smarter Balanced Mathematics Test Scores for Grade Eight

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number Tested** | **Correlation** |
| All students | 424,176 | 0.82 |
| Male | 217,339 | 0.82 |
| Female | 206,462 | 0.81 |
| Nonbinary | 375 | 0.80 |
| EL | 53,836 | 0.53 |
| English only | 241,072 | 0.81 |
| RFEP | 111,049 | 0.78 |
| IFEP | 18,130 | 0.82 |
| ADEL | 0 | N/A |
| To be determined | 55 | 0.73 |
| English proficiency unknown | 34 | 0.78 |
| Economically disadvantaged | 275,930 | 0.77 |
| Not economically disadvantaged | 148,246 | 0.82 |
| American Indian or Alaska Native (All) | 1,785 | 0.75 |
| Asian (All) | 42,364 | 0.83 |
| Native Hawaiian or Other Pacific Islander (All) | 1,837 | 0.79 |
| Filipino (All) | 10,032 | 0.81 |
| Hispanic or Latino (All) | 240,524 | 0.76 |
| Black or African American (All) | 21,164 | 0.74 |
| White (All) | 84,796 | 0.80 |
| Two or more races (All) | 21,674 | 0.83 |
| Disability | 49,905 | 0.71 |
| No disability | 374,271 | 0.81 |
| Migrant education | 3,326 | 0.73 |
| Not migrant education | 420,850 | 0.82 |
| Armed forces family member | 5,940 | 0.81 |
| Not armed forces family member | 418,236 | 0.82 |
| Homeless | 13,596 | 0.74 |
| Not homeless | 410,580 | 0.82 |
| Foster youth | 1,390 | 0.68 |
| Not foster youth | 422,786 | 0.82 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 521 | 0.75 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 1,264 | 0.73 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 24,547 | 0.80 |
| Asian (Primary ethnicity—Economically disadvantaged) | 17,817 | 0.83 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 515 | 0.82 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 1,322 | 0.74 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 5,903 | 0.80 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 4,129 | 0.80 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 48,116 | 0.79 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 192,408 | 0.74 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 4,853 | 0.77 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 16,311 | 0.71 |
| White (Primary ethnicity—Not economically disadvantaged) | 51,457 | 0.78 |
| White (Primary ethnicity—Economically disadvantaged) | 33,339 | 0.78 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 12,334 | 0.81 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 9,340 | 0.79 |

Table 8.K.5 Correlations with Smarter Balanced ELA Test Scores for Grade Eleven

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number Tested** | **Correlation** |
| All students | 327,475 | 0.78 |
| Male | 167,578 | 0.79 |
| Female | 159,237 | 0.78 |
| Nonbinary | 660 | 0.75 |
| EL | 33,520 | 0.54 |
| English only | 176,921 | 0.78 |
| RFEP | 101,432 | 0.73 |
| IFEP | 15,558 | 0.76 |
| ADEL | 0 | N/A |
| To be determined | 20 | 0.82 |
| English proficiency unknown | 24 | 0.85 |
| Economically disadvantaged | 203,724 | 0.76 |
| Not economically disadvantaged | 123,751 | 0.78 |
| American Indian or Alaska Native (All) | 1,413 | 0.75 |
| Asian (All) | 32,960 | 0.79 |
| Native Hawaiian or Other Pacific Islander (All) | 1,385 | 0.74 |
| Filipino (All) | 8,546 | 0.75 |
| Hispanic or Latino (All) | 185,014 | 0.75 |
| Black or African American (All) | 14,449 | 0.74 |
| White (All) | 68,714 | 0.77 |
| Two or more races (All) | 14,994 | 0.79 |
| Disability | 34,881 | 0.70 |
| No disability | 292,594 | 0.77 |
| Migrant education | 2,194 | 0.75 |
| Not migrant education | 325,281 | 0.78 |
| Armed forces family member | 4,927 | 0.77 |
| Not armed forces family member | 322,548 | 0.78 |
| Homeless | 9,769 | 0.75 |
| Not homeless | 317,706 | 0.78 |
| Foster youth | 1,090 | 0.72 |
| Not foster youth | 326,385 | 0.78 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 499 | 0.76 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 914 | 0.72 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 18,587 | 0.76 |
| Asian (Primary ethnicity—Economically disadvantaged) | 14,373 | 0.79 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 500 | 0.75 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 885 | 0.73 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 5,134 | 0.75 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 3,412 | 0.75 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 40,681 | 0.76 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 144,333 | 0.74 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 4,147 | 0.75 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 10,302 | 0.72 |
| White (Primary ethnicity—Not economically disadvantaged) | 45,108 | 0.76 |
| White (Primary ethnicity—Economically disadvantaged) | 23,606 | 0.76 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 9,095 | 0.77 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 5,899 | 0.77 |

Table 8.K.6 Correlations with Smarter Balanced Mathematics Test Scores for Grade Eleven

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number Tested** | **Correlation** |
| All students | 327,640 | 0.78 |
| Male | 167,712 | 0.79 |
| Female | 159,270 | 0.77 |
| Nonbinary | 658 | 0.76 |
| EL | 34,336 | 0.49 |
| English only | 176,467 | 0.78 |
| RFEP | 101,244 | 0.74 |
| IFEP | 15,549 | 0.78 |
| ADEL | 0 | N/A |
| To be determined | 21 | 0.65 |
| English proficiency unknown | 23 | 0.74 |
| Economically disadvantaged | 203,805 | 0.73 |
| Not economically disadvantaged | 123,835 | 0.79 |
| American Indian or Alaska Native (All) | 1,384 | 0.72 |
| Asian (All) | 33,007 | 0.79 |
| Native Hawaiian or Other Pacific Islander (All) | 1,379 | 0.72 |
| Filipino (All) | 8,553 | 0.76 |
| Hispanic or Latino (All) | 185,299 | 0.72 |
| Black or African American (All) | 14,392 | 0.71 |
| White (All) | 68,657 | 0.77 |
| Two or more races (All) | 14,969 | 0.80 |
| Disability | 34,736 | 0.64 |
| No disability | 292,904 | 0.77 |
| Migrant education | 2,218 | 0.70 |
| Not migrant education | 325,422 | 0.78 |
| Armed forces family member | 4,934 | 0.76 |
| Not armed forces family member | 322,706 | 0.78 |
| Homeless | 9,899 | 0.70 |
| Not homeless | 317,741 | 0.78 |
| Foster youth | 1,084 | 0.68 |
| Not foster youth | 326,556 | 0.78 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 494 | 0.75 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 890 | 0.67 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 18,595 | 0.77 |
| Asian (Primary ethnicity—Economically disadvantaged) | 14,412 | 0.78 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 499 | 0.74 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 880 | 0.69 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 5,145 | 0.75 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 3,408 | 0.75 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 40,857 | 0.75 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 144,442 | 0.70 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 4,130 | 0.74 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 10,262 | 0.67 |
| White (Primary ethnicity—Not economically disadvantaged) | 45,038 | 0.76 |
| White (Primary ethnicity—Economically disadvantaged) | 23,619 | 0.73 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 9,077 | 0.79 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 5,892 | 0.76 |

### Appendix 8.L: Correlations Between CAST Domain Scores and Total Test Scores and Smarter Balanced Composite Claim Scores and Total Test Scores

**Note:** For table 8.L.1 through table 8.L.6, number tested = students who received a score on both a CAST domain or total test and a Smarter Balanced claim or total test, SB = Smarter Balanced, and ELA = English language arts/literacy.

Table 8.L.1 Correlations Between CAST Domain and Total Test Scores and Smarter Balanced ELA Composite Claim and Total Test Scores, Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SB Composite Claims** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** | **Total CAST Score** |
| Reading and Listening—Number Tested | 416,451 | 416,459 | 416,442 | 417,114 |
| Reading and Listening—Correlation | 0.73 | 0.73 | 0.70 | 0.78 |
| Writing and Research—Number Tested | 416,445 | 416,453 | 416,436 | 417,106 |
| Writing and Research—Correlation | 0.74 | 0.75 | 0.71 | 0.79 |
| **Total SB ELA—Number Tested:** | **416,471** | **416,479** | **416,463** | **417,138** |
| **Total SB ELA—Correlation:** | **0.79** | **0.79** | **0.76** | **0.84** |

Table 8.L.2 Correlations Between CAST Domain and Total Test Scores and Smarter Balanced Mathematics Composite Claim and Total Test Scores, Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SB Composite Claims** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** | **Total CAST Score** |
| Concepts and Procedures—Number Tested | 420,614 | 420,629 | 420,603 | 421,276 |
| Concepts and Procedures—Correlation | 0.71 | 0.69 | 0.68 | 0.74 |
| Problem Solving, Communicating Reasoning, and Modeling and Data Analysis—Number Tested | 420,627 | 420,642 | 420,615 | 421,290 |
| Problem Solving, Communicating Reasoning, and Modeling and Data Analysis—Correlation | 0.75 | 0.73 | 0.71 | 0.79 |
| **Total SB Mathematics—Number Tested:** | **420,627** | **420,642** | **420,615** | **421,290** |
| **Total SB Mathematics—Correlation:** | **0.77** | **0.75** | **0.74** | **0.81** |

Table 8.L.3 Correlations Between CAST Domain and Total Test Scores and Smarter Balanced ELA Composite Claim and Total Test Scores, Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SB Composite Claims** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** | **Total CAST Score** |
| Reading and Listening—Number Tested | 420,624 | 420,571 | 420,055 | 421,539 |
| Reading and Listening—Correlation | 0.66 | 0.74 | 0.71 | 0.76 |
| Writing and Research—Number Tested | 420,686 | 420,633 | 420,116 | 421,607 |
| Writing and Research—Correlation | 0.64 | 0.74 | 0.71 | 0.76 |
| **Total SB ELA—Number Tested:** | **420,793** | **420,737** | **420,216** | **421,734** |
| **Total SB ELA—Correlation:** | **0.69** | **0.79** | **0.76** | **0.81** |

Table 8.L.4 Correlations Between CAST Domain and Total Test Scores and Smarter Balanced Mathematics Composite Claim and Total Test Scores, Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SB Composite Claims** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** | **Total CAST Score** |
| Concepts and Procedures—Number Tested | 423,175 | 423,123 | 422,624 | 424,083 |
| Concepts and Procedures—Correlation | 0.65 | 0.72 | 0.72 | 0.76 |
| Problem Solving, Communicating Reasoning, and Modeling and Data Analysis—Number Tested | 423,252 | 423,198 | 422,693 | 424,174 |
| Problem Solving, Communicating Reasoning, and Modeling and Data Analysis—Correlation | 0.67 | 0.74 | 0.74 | 0.78 |
| **Total SB Mathematics—Number Tested:** | **423,254** | **423,200** | **422,695** | **424,176** |
| **Total SB Mathematics—Correlation:** | **0.70** | **0.77** | **0.77** | **0.82** |

Table 8.L.5 Correlations Between CAST Domain and Total Test Scores and Smarter Balanced ELA Composite Claim and Total Test Scores, Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SB Composite Claims** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** | **Total CAST Score** |
| Reading and Listening—Number Tested | 326,637 | 326,456 | 326,011 | 327,307 |
| Reading and Listening—Correlation | 0.67 | 0.68 | 0.66 | 0.72 |
| Writing and Research—Number Tested | 326,637 | 326,457 | 326,011 | 327,309 |
| Writing and Research—Correlation | 0.67 | 0.71 | 0.69 | 0.74 |
| **Total SB ELA—Number Tested:** | **326,781** | **326,597** | **326,139** | **327,475** |
| **Total SB ELA—Correlation:** | **0.72** | **0.74** | **0.72** | **0.78** |

Table 8.L.6 Correlations Between CAST Domain and Total Test Scores and Smarter Balanced Mathematics Composite Claim and Total Test Scores, Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SB Composite Claims** | **Earth and Space Sciences** | **Life Sciences** | **Physical Sciences** | **Total CAST Score** |
| Concepts and Procedures—Number Tested | 326,955 | 326,766 | 326,305 | 327,613 |
| Concepts and Procedures—Correlation | 0.67 | 0.67 | 0.69 | 0.73 |
| Problem Solving, Communicating Reasoning, and Modeling and Data Analysis—Number Tested | 326,969 | 326,780 | 326,317 | 327,632 |
| Problem Solving, Communicating Reasoning, and Modeling and Data Analysis—Correlation | 0.68 | 0.67 | 0.69 | 0.73 |
| **Total SB Mathematics—Number Tested:** | **326,977** | **326,787** | **326,324** | **327,640** |
| **Total SB Mathematics—Correlation:** | **0.72** | **0.71** | **0.73** | **0.78** |

## Quality-Control Procedures

The California Department of Education (CDE) and ETS implemented rigorous quality-control procedures throughout the assessment development, administration, scoring, analyses, and reporting processes for CAST. As part of this effort, ETS staff worked with its Office of Professional Standards Compliance, which publishes and maintains the *ETS Standards for Quality and Fairness* (ETS, 2014). These *Standards* support the goals of delivering technically sound, fair, and useful products and services; and assisting the public and auditors evaluating those products and services. Quality-control procedures are outlined in this chapter.

### Quality Control of Item Development

The ETS goal is to provide the best standards-based and innovative items for the CAST. Items developed for CAST were subject to an extensive item review process. The item writers responsible for developing CAST items were trained in the California Assessment of Student Performance and Progress (CAASPP) and ETS policies on quality control of item content, bias and sensitivity guidelines, as well as guidelines for accessibility, to ensure that the items allow the widest possible range of students to demonstrate their abilities.

Once a draft item was accepted for authoring—that is, once it was entered into the ETS item bank and formatted for use in an assessment—ETS employed a series of internal reviews and an initial CDE review. These reviews used established criteria and specifications to judge the quality of an item’s content and ensured that each item measured what it was intended to measure. These reviews also examined the overall quality of the test items before presentation to the CDE and item reviewers. To finish the process, a group of California educators reviewed the items for accessibility, bias and sensitivity, and content, and made recommendations for item enhancement. The details on quality control of item development are described in section [*3.3 ETS Item Review Process*](#_ETS_Item_Review_1).

During administrations of CAST, when sufficient student response data on each item became available, ETS Psychometric Analysis & Research (PAR) staff conducted item analyses and a key check to examine whether the items performed as expected. ETS psychometric staff conducted a thorough evaluation of all item statistics using the statistical criteria described in subsection [*8.2.7 Summary of Classical Item Analyses Flagging Criteria*](#_Summary_of_Classical_1) to flag items that were potentially problematic because of poor item performance, content issues, item bias, or accessibility challenges. Flagged items were then reviewed by ETS Assessment Development (AD) staff, the CDE, and California educators to determine whether issues existed.

### Quality Control of Test Assembly and Delivery

The assembly of all test forms must conform to blueprints that represent a set of constraints and specifications. ETS conducted multiple levels of quality assurance (QA) checks on each assembled CAST form to ensure it met the form-building specifications. Both ETS AD and PAR staff reviewed and signed off on the accuracy of forms before the test forms were posted for CDE review. Detailed information related to test assembly can be found in [*Chapter 4: Test Assembly*](#_Test_Assembly).

In particular, the assembly of all test forms went through a certification process that involved various checks, including verifying that

* all item answers in the key were correctly identified and documented in the scoring system;
* items were scored correctly in the item bank and incorrect responses were scored as incorrect;
* all items assessed the intended standard;
* all content in the item was correct with the exception of distractors, which are intended to be incorrect;
* all items met the statistical criteria, to the extent possible;
* distractors were plausible;
* multiple-choice item options were parallel in structure;
* language was grade-level appropriate;
* no more than three multiple-choice items in a row had the same key;
* all graphics were correct (copyright, spelling, relevance, etc.);
* there were no unintended mechanical errors in grammar, spelling, punctuation, and the like; and
* items adhered to the approved style guide.

Reviews were also conducted for functionality and sequencing during the user acceptance testing (UAT) process to ensure all items functioned as expected. Three cycles of UAT were conducted: the first by the Test Delivery System (TDS) vendor, the second by ETS, and the third by the CDE. CDE staff made a final quality check to ensure that all issues identified during UAT were resolved before the release of the operational assessment.

### Quality Control of Test Materials

Brief descriptions of the types of materials used for and during testing appear in the following subsections.

#### Developing Assessments

##### Computer-based Assessments

The steps taken to develop and ensure the quality of the computer-based assessments are described in [*Chapter 3: Item Development and Review*](#_Item_Development_and) and [*Chapter 4:* *Test Assembly*](#_Test_Assembly).

##### Paper–Pencil Forms

Test forms and response booklets were developed and reviewed by ETS staff to ensure that materials met quality standards. Each document was reviewed for accuracy, completeness, and alignment with supporting materials. Print-ready PDFs underwent a stringent quality-control process to ensure that there was adequate space for student response.

#### Test Administration Manuals

ETS staff verified that test instruction manuals accurately matched the test materials and testing processes. Editors reviewed each document for spelling, grammar, accuracy, and adherence to CDE style. Each document was approved by the CDE before being published to the CAASPP & English Language Proficiency Assessments for California (ELPAC) Website. Only nonsecure documents were posted to this website. Secure materials, such as the *CAASPP Directions for Administration* (*DFAs*), were made available to designated local educational agency (LEA) staff through the Test Operations Management System (TOMS), which required a secure logon.

The manuals used in the administration of the CAASPP are listed in subsection [*5.2.4 Instructions for Test Administration*](#_Instructions_for_Test_1).

#### Collecting Test Materials

ETS processes ensure the security of assessments delivered using a variety of test modes and delivery methods.

##### Computer-based Assessments

During the 2023–24 CAASPP administration, there were no test materials to be collected as a result of computer-based testing.

##### Paper–Pencil Forms

Once the paper–pencil tests (PPTs) were administered at test sites whose LEAs had received prior approval from the CDE, LEAs were instructed to enter student responses into the CAASPP Data Entry Interface (DEI). LEAs were provided with instructions for secure destruction of materials in the *CAASPP Paper–Pencil Testing Test Administration Manual* (CDE, 2024).

#### Processing Test Materials

The ways in which materials associated with student testing were processed are described in subsequent subsections.

##### Computer-based Assessments

Computer-based assessments submitted by students were transmitted from Cambium Assessment, Inc. (CAI) to ETS each day. Each system checked for the completeness of the student record and stopped records that were identified as having an error. (For example, the system would identify a test part that was missing a content registration ID, which is a unique identifier that matches the student’s available test opportunities.)

##### Paper–Pencil Forms

Once student responses were entered into the DEI, and within five working days after the last day of each test administration period, LEAs returned all scorable and nonscorable materials. The LEAs packed all materials into cartons, applied the labels, and then numbered the cartons prior to returning the materials to ETS by means of their assigned carrier.

### Quality Control of Test Administration

The quality of test administration for CAST was monitored and controlled through several strategies.

A fully supported Outreach team that includes California Technical Assistance Center phone support and Success Agents supported all LEAs in the administration of the CAASPP. In addition to providing guidance and answering questions, the Outreach team regularly conducted campaigns on particular administration topics to ensure all LEAs understood correct test administration procedures. Outreach was guided by individuals who managed communications to LEAs; provided regional and web-based trainings; and hosted a website, [the](https://www.caaspp.org/) CAASPP & ELPAC Website, that housed a full range of manuals, videos, and other instructional and support materials.

The quality of test administration was further managed through comprehensive rules and guidelines for maintaining the security and standardization of the CAASPP. LEAs received training on these topics and were provided tools for reporting security incidents and resolving testing discrepancies for specific testing sessions.

The ETS Office of Testing Integrity (OTI) reinforced the quality-control procedures for test administration, providing QA services for all testing programs managed by ETS. The detailed procedures the OTI developed and applied in quality control are described in subsection [*5.6.1 The ETS Office of Testing Integrity*](#_The_ETS_Office_1).

### Quality Control of Scoring

ETS conforms to high standards of quality and fairness when scoring assessments and reporting scores. These standards dictate that ETS provides accurate and understandable assessment results to the intended recipients. It is also the ETS 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 demographic student groups.

#### Machine-Scoring Procedures

To ensure valid item-level scoring for CAST, quality-control procedures were employed by CAI, the CAASPP subcontractor responsible for providing the TDS and scoring machine-scorable items. CAI staff independently reviewed all CAST forms by producing sample results for assessments. The sample results were compared with the answer keys for each form to confirm the accuracy of scoring keys. The scores for all applicable items were recorded. A final comparison of the test map to each computer-based form as configured in the UAT environment ensured that no changes to the form were introduced prior to operational deployment.

A real-time, quality-monitoring component was built into the TDS. After an assessment was administered to a student, the TDS passed the resulting data to the QA system. QA conducted a series of data integrity checks, ensuring, for example, that the record for each assessment contained information for each item, keys for multiple-choice items, score points in each item, and the total number of operational items. In addition, QA also checked to ensure that the test record contained no data from items that might have been invalidated.

Data passed directly from the Quality Monitoring System to the database of record, which served as the repository for all test information, and from which all test information was pulled and transmitted to ETS in a predetermined results format.

#### Human Scoring

For human scoring, ETS employed multiple quality controls, including

* raters being required to successfully pass calibration, described in subsection [*7.1.1.2.6 Training for Raters*](#_Training_for_Raters), prior to beginning scoring at each grade level;
* scoring leaders conducting backreads during each scoring shift;
* ETS reviewing statistics on validity papers; and
* ETS reviewing interrater reliability statistics.

Refer to subsection[*7.1.1.2 Human Scoring Development*](#_Human_Scoring_Development) for the topics; refer to subsections [*7.1.1.2.8 Scoring Monitoring and Quality Management*](#_Scoring_Monitoring_and) and [*7.1.1.2.10 Validity Responses and Sets*](#_Validity_Responses_and) for more specific details on these tools used for quality control of human scoring.

#### Artificial Intelligence Scoring Verification

During the 2023–24 CAST administration, to ensure the quality of artificial intelligence (AI) scoring, ETS maintained a QA system where a random sample of human ratings was obtained and used for rater agreement analyses. All AI-scored items were subject to a second scoring by human raters. Refer to subsection [*7.1.1.2.9 Interrater Reliability for Operational Items*](#_Interrater_Reliability_for) for more information regarding the scoring process.

#### Development of Scoring Specifications

A number of measures were taken to ascertain that the scoring keys were applied to the student responses as intended and the student scores were computed accurately. ETS built and reviewed the scoring system models based on the reporting specifications approved by the CDE. These specifications contain detailed scoring procedures, along with the procedures for determining whether a student has attempted an assessment and whether that student’s response data should be included in the statistical analyses and calculations for computing summary data.

Prior to the test administration, ETS AD staff reviewed and verified the keys and scoring rubrics for each item. Then, these keys andrubrics were provided to CAI for implementing machine scoring of the selected-response items. Item responses to be human-scored were sent electronically to the ETS Online Network for Evaluation for scoring by trained, qualified raters. In addition, the student’s original response string was stored for data verification and auditing purposes. Standard quality inspections were performed on all data files, including the evaluation of each student data record for correctness and completeness. Student results were kept confidential and secure at all times.

ETS scoring specifications for CAST were completed, approved, and checked well in advance of the receipt of student response data.

#### Paper–Pencil Scoring

If an LEA was approved to administer the PPT version of CAST, student responses were entered into the DEI and scored electronically and by a rater, depending on the item type.

### Quality Control of Psychometric Processes

#### Scoring Verification

ETS developed two independent and parallel scoring structures to produce students’ scores: the Enterprise Score Key Management (eSKM) scoring system, which collected, scored, and delivered individual students’ scores to the ETS reporting system; and then the ETS PAR team computed individual student scores based on the same scoring specifications as described in subsection [*9.5.4 Development of Scoring Specifications*](#_Development_of_Scoring). The scores from the two sources were then compared for internal quality control. Any differences in the scores were discussed and resolved. All scores complied with the ETS scoring specifications and passed the parallel scoring process. This ensured the quality and accuracy of scoring and supported the transfer of scores into TOMS, the database of the student records scoring.

#### Psychometric Analyses

The psychometric procedures for CAST were developed, reviewed, and approved prior to the receipt of student response data. The ETS psychometric team also developed specifications for each of the psychometric analyses performed. These specifications contain detailed descriptions of the analysis steps such as sample inclusion, analyses methods, and special handling of the data.

All psychometric analyses conducted at ETS underwent comprehensive quality checks by a team of psychometricians and data analysts. Detailed checklists and psychometric specifications were developed by members of the team for each of the statistical procedures performed on CAST results data, including item analyses, differential item functioning analyses, item response theory (IRT) calibration, equating, and scaling.

Detailed checklists were developed by members of the team for each of the statistical procedures. Classical item analyses were performed to evaluate the performance of the operational items. Classical item statistics included item difficulty and correlations between item scores and total scores. Items that were flagged for questionable statistical attributes were sent to ETS AD staff for review; their comments were then reviewed by the psychometricians before the review by the CDE. The ETS AD and PAR teams worked together to evaluate and make recommendations to the CDE about any problematic items that should be removed from IRT calibration.

IRT calibration of field test items included checks to ascertain that the input files were established accurately. Checks were also made on the number of items, number of students with valid scores, IRT item difficulty and discrimination estimates, standard errors for the item difficulty estimates, and the equating and scaling process. Two psychometricians conducted parallel calibration processing and compared the results to check for any inconsistency. Psychometricians also performed detailed reviews of relevant statistics to determine whether the chosen IRT model fits the data. ETS then presented and reviewed the calibration results with the CDE for approval.

Once raw-to-scale-score conversion tables for each form were generated, psychometricians carried out quality-control checks on each scoring table to verify

* all possible raw scores for each form were included in the tables;
* the lowest obtainable scale score and the highest obtainable scale score matched the specifications for each grade level, respectively; and
* the threshold score for the achievement level was correctly identified.

After all quality-control steps were completed and any differences were resolved, one final inspection of scoring tables was conducted prior to uploading the tables to eSKM for score reporting.

### Quality Control of Reporting

To ensure the quality of CAST results, for both individual student and summary reports, three general areas were evaluated:

1. Comparison of report formats with input sources from the CDE-approved samples
2. Validation of the report data through quality-control checks performed by the ETS Data Quality Services and Center of Reporting & Scoring Services teams, as well as running of all Student Score Reports (SSRs) through the ETS patented QC Interrogator software, which compares elements of the SSR to acceptable values to identify errors and is used in conjunction with human review to detect errors on every score report batch as part of quality-control procedures
3. Proofreading of the quality-control and production reports by the CDE and ETS prior to making reports available to the LEA for download in TOMS and the California Educator Reporting System as well as via the LEA student information system

All reports were required to include a single, accurate LEA code, an LEA name, and a school name. All elements conformed to the CDE official county/district/school (CDS) code and naming records. From the start of processing through scoring and reporting, the CDS Master File was used to verify and confirm accurate codes and names. The CDE provided a revised LEA Master File to ETS throughout the year as updates became available.

After the reports were validated in accordance with CDE requirements, a set of reports representing all possible grade levels, content areas, and reporting outcomes was provided to the CDE and ETS for review and approval. Electronic reports were sent on the actual report template to the CDE. The CDE and ETS reviewed and approved the reports after a thorough examination.

Upon CDE approval of the reports generated for the quality-control LEAs, ETS proceeded with the first batch of report production. The reviewed set of reports incorporated CDE-selected LEAs and provided the final check prior to generating all reports and making them available electronically for download in TOMS and for student information systems through an application programming interface.

#### Exclusion of Student Scores from Summary Reports

ETS provided the CDE with reporting specifications that documented when to exclude student scores from summary reports. These specifications included the logic for handling submitted assessments that, for example, indicated the student tested but responded to no items, was absent, was not tested because of parent/guardian request, or did not complete the assessment because of illness. The methods for handling other anomalies were also covered in the specifications. These anomalies are described in more detail in [*7.4.2 Special Cases*](#_Special_Cases_1).

### Quality Control of End-to-End Testing

ETS conducted end-to-end testing prior to the start of the test administration. The purpose of this testing is to verify that all systems, processes, and resources were ready for the operational administration. Once released from processing, the test results were sent through the system for scoring and reporting. SSRs were created, along with data files for subject-matter experts in the teams to review and verify.

#### Computer-based Assessments

ETS employed a number of strategies to verify ongoing systems performance, including monitoring of system availability and system usage. Time was allotted for UAT to confirm that the systems met requirements and to make identified corrections before final deployment. To accomplish system acceptance and sign-off, ETS deployed systems to a staging area, which mirrors the final production environment, for operational testing and UAT. Final approval by the CDE triggered final deployment of the system.

To begin the quality-control process for end-to-end testing of the administration, the ETS program and resolutions teams prepared by entering responses in computer-based assessments for all grade levels and the high school grade band. These responses were entered for fictitious students in selected schools and across several LEAs. Each student’s assessment was completed with responses that were all correct, all incorrect, and combinations of correct and incorrect. These response combinations were the expected results across achievement levels and score ranges. The responses were sent for processing, including for system quality control of computer-based assessments.

Once released from processing, the test results were sent through the system for scoring and reporting. SSRs were created, along with data files for subject-matter experts in the teams to review and verify. Individual SSRs were generated on the basis of the fictitious students when 100 percent quality control was demonstrated by the ETS Resolution staff.

#### Paper–Pencil Tests

The DEI underwent UAT to ensure that the correct test items were available for a grade-level assessment in the DEI. Then, during testing, information technology personnel monitored daily feeds to ensure the completeness and timeliness of records sent for hand scoring.

The processes followed to test the DEI from end to end are described in the previous subsection, [*9.8.1 Computer-based Assessments*](#_Computer-based_Assessments).

### References

California Department of Education. (2024). *CAASPP paper–pencil testing test administration manual*. Sacramento, CA: California Department of Education.

Educational Testing Service. (2014). *ETS Standards for Quality and Fairness*. Princeton, NJ: Educational Testing Service.

## Student Survey

This chapter describes the development and administration of the survey questions presented to students during the 2023–24 California Science Test (CAST) operational test administration. The summary of findings and results of analyses of the survey data are included.

Student survey questions were developed by the California Department of Education to gather information about how the science content on the CAST compared to the science content presented in the classroom. The wording and options for student survey questions remained unchanged from those used during the 2022–23 test administration.

### Student Survey Questions

The survey questionnaire was administered after students completed the CAST. There were four survey questions, each with four response options:

1. How many topics on the test were taught in your science classes?

* **All** of the topics on the test were taught in my science classes.
* **Most** of the topics on the test were taught in my science classes, but not all of them.
* **Some** of the topics on the test were taught in my science classes, but not most of them.
* **None** of the topics on the test were taught in my science classes.

1. Were any test questions different from the types of questions you see in your science class?

* **All** of the questions on the test were different.
* **Most** of the questions on the test were different.
* **Some** of the questions on the test were different.
* **None** of the questions on the test were different.

1. How hard were questions on this test compared to questions you see in your science class?

* They were **harder** than most questions in my science classes.
* They were **about as hard** as the questions in my science classes.
* They were **easier** than most questions in my science classes.

1. Do you think you will be enrolling in any more science classes in high school?

* Yes, I **will be** enrolling in additional high school science classes.
* No, I will **not** be taking any additional high school science classes.

Questions one through three were for all students. Question four was applicable only to high school students in grades ten and eleven. Refer to [appendix 10.A](#_Appendix_10.A:_Results_1) for details on the options for each question and student response frequencies.

The questions were available in braille for students who used the braille accommodation.

### Student Survey Results

Table 10.A.1 through table 10.A.3 show the results for the three survey questions for grade five students. Table 10.A.4 through table 10.A.6 show the survey results for grade eight students. Table 10.A.7 through table 10.A.21 show the results for the survey questions by grade level and for high school students overall. Table 10.A.21 shows results only for grade ten and grade eleven students, since the fourth survey question does not apply to grade twelve students.

For the first survey question, “How many topics on the test were taught in your science classes?,” 55.2 percent of the students in grade five, 65.4 percent in grade eight, and 47.4 percent in the high school grade band (that is, grade ten, eleven, or twelve) agreed that most or all of the topics were taught in their science classes. Among high school students, this response option was selected by 43.0 percent of grade ten students, 47.5 percent of grade eleven students, and 47.8 percent of grade twelve students.

For the second survey question, “Were any test questions different from the types of questions you see in your science class?,” 44.8 percent of the students in grade five, 41.5 percent in grade eight, and 52.7 percent in the high school grade band agreed that most or all of the questions on the test were different, that is, chose the first or the second option. Among high school students, these two response options were selected by 54.7 percent of grade ten students, 52.7 percent of grade eleven students, and 52.0 percent of grade twelve students. Across all grade levels and the high school grade band, less than 6 percent of students responded that none of the questions on the test were different.

For the third survey question, “How hard were questions on this test compared to questions you see in your science class?,” 46.2 percent to 57.7 percent of students selected the second option, indicating that they felt the test questions were about as hard as the questions in their science classes across all grade levels and the high school grade band. Additionally, more students indicated that the test questions on the test were harder than most questions in their science classes (the first option) than those indicating that the test questions were easier (the third option) across all grade levels and the high school grade band.

For the fourth survey question, “Do you think you will be enrolling in any more science classes in high school?,” 59.4 percent of grade ten students and 47.2 percent of grade eleven students responded that they would enroll in more science classes during high school.

Table 10.A.22 presents the correlations between student survey responses and the CAST scale scores. The negative correlations between student survey responses to the fourth question and their scale scores for grade ten and grade eleven students indicate that students who performed well tended to be planning to enroll in additional high school science classes. The correlations between student survey responses to the first three questions and their scale scores across all grade levels ranged from −0.02 to 0.29, indicating there was no clear relationship between student responses on those survey questions and their performance on the CAST.

### Appendix 10.A: Results of the Student Survey

**Note:** “SD” is standard deviation.

Table 10.A. Scale Score Distribution by Responses to Question 1 for Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **How many topics on the test were taught in your science classes?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| **All** of the topics on the test were taught in my science classes. | 51,040 | 12.1 | 196 | 24.4 |
| **Most** of the topics on the test were taught in my science classes, but not all of them. | 181,366 | 43.1 | 204 | 22.6 |
| **Some** of the topics on the test were taught in my science classes, but not most of them. | 166,042 | 39.4 | 202 | 20.8 |
| **None** of the topics on the test were taught in my science classes. | 22,791 | 5.4 | 197 | 20.9 |
| **Total:** | **421,239** | **100.0** | **202** | **22.2** |

Table 10.A. Scale Score Distribution by Responses to Question 2 for Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Were any test questions different from the types of questions you see in your science class?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| **All** of the questions on the test were different. | 48,799 | 11.6 | 191 | 21.7 |
| **Most** of the questions on the test were different. | 139,892 | 33.2 | 201 | 22.0 |
| **Some** of the questions on the test were different. | 214,160 | 50.8 | 205 | 21.5 |
| **None** of the questions on the test were different. | 18,384 | 4.4 | 195 | 23.0 |
| **Total:** | **421,235** | **100.0** | **202** | **22.2** |

Table 10.A. Scale Score Distribution by Responses to Question 3 for Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **How hard were questions on this test compared to questions you see in your science class?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| They were **harder** than most questions in my science classes. | 137,729 | 32.7 | 200 | 22.0 |
| They were **about as hard** as the questions in my science classes. | 243,027 | 57.7 | 203 | 21.9 |
| They were **easier** than most questions in my science classes. | 40,476 | 9.6 | 198 | 23.9 |
| **Total:** | **421,232** | **100.0** | **202** | **22.2** |

Table 10.A. Scale Score Distribution by Responses to Question 1 for Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **How many topics on the test were taught in your science classes?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| **All** of the topics on the test were taught in my science classes. | 69,885 | 16.5 | 405 | 24.1 |
| **Most** of the topics on the test were taught in my science classes, but not all of them. | 207,887 | 48.9 | 403 | 22.2 |
| **Some** of the topics on the test were taught in my science classes, but not most of them. | 135,004 | 31.8 | 397 | 19.9 |
| **None** of the topics on the test were taught in my science classes. | 12,011 | 2.8 | 388 | 17.8 |
| **Total:** | **424,787** | **100.0** | **401** | **22.1** |

Table 10.A. Scale Score Distribution by Responses to Question 2 for Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Were any test questions different from the types of questions you see in your science class?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| **All** of the questions on the test were different. | 39,323 | 9.3 | 389 | 19.2 |
| **Most** of the questions on the test were different. | 136,878 | 32.2 | 399 | 21.2 |
| **Some** of the questions on the test were different. | 229,320 | 54.0 | 404 | 22.0 |
| **None** of the questions on the test were different. | 19,261 | 4.5 | 399 | 24.7 |
| **Total:** | **424,782** | **100.0** | **401** | **22.1** |

Table 10.A. Scale Score Distribution by Responses to Question 3 for Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **How hard were questions on this test compared to questions you see in your science class?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| They were **harder** than most questions in my science classes. | 173,176 | 40.8 | 398 | 20.7 |
| They were **about as hard** as the questions in my science classes. | 220,865 | 52.0 | 403 | 22.3 |
| They were **easier** than most questions in my science classes. | 30,736 | 7.2 | 399 | 26.3 |
| **Total:** | **424,777** | **100.0** | **401** | **22.1** |

Table 10.A. Scale Score Distribution by Responses to Question 1 for Grade Ten

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **How many topics on the test were taught in your science classes?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| **All** of the topics on the test were taught in my science classes. | 1,672 | 6.8 | 593 | 21.6 |
| **Most** of the topics on the test were taught in my science classes, but not all of them. | 8,913 | 36.2 | 603 | 22.2 |
| **Some** of the topics on the test were taught in my science classes, but not most of them. | 12,544 | 50.9 | 598 | 18.5 |
| **None** of the topics on the test were taught in my science classes. | 1,498 | 6.1 | 587 | 14.4 |
| **Total:** | **24,627** | **100.0** | **599** | **20.3** |

Table 10.A. Scale Score Distribution by Responses to Question 2 for Grade Ten

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Were any test questions different from the types of questions you see in your science class?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| **All** of the questions on the test were different. | 3,184 | 12.9 | 590 | 17.3 |
| **Most** of the questions on the test were different. | 10,293 | 41.8 | 600 | 19.6 |
| **Some** of the questions on the test were different. | 10,339 | 42.0 | 602 | 20.8 |
| **None** of the questions on the test were different. | 811 | 3.3 | 589 | 19.2 |
| **Total:** | **24,627** | **100.0** | **599** | **20.3** |

Table 10.A. Scale Score Distribution by Responses to Question 3 for Grade Ten

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **How hard were questions on this test compared to questions you see in your science class?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| They were **harder** than most questions in my science classes. | 10,986 | 44.6 | 595 | 18.0 |
| They were **about as hard** as the questions in my science classes. | 11,436 | 46.4 | 601 | 20.5 |
| They were **easier** than most questions in my science classes. | 2,205 | 9.0 | 608 | 25.2 |
| **Total:** | **24,627** | **100.0** | **599** | **20.3** |

Table 10.A. Scale Score Distribution by Responses to Question 4 for Grade Ten

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Do you think you will be enrolling in any more science classes in high school?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| Yes, I **will be** enrolling in additional high school science classes. | 14,638 | 59.4 | 604 | 21.1 |
| No, I will **not** be taking any additional high school science classes. | 9,988 | 40.6 | 592 | 16.7 |
| **Total:** | **24,626** | **100.0** | **599** | **20.3** |

Table 10.A. Scale Score Distribution by Responses to Question 1 for Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **How many topics on the test were taught in your science classes?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| **All** of the topics on the test were taught in my science classes. | 28,498 | 8.6 | 601 | 25.6 |
| **Most** of the topics on the test were taught in my science classes, but not all of them. | 128,303 | 38.9 | 607 | 23.1 |
| **Some** of the topics on the test were taught in my science classes, but not most of them. | 154,081 | 46.8 | 600 | 19.2 |
| **None** of the topics on the test were taught in my science classes. | 18,690 | 5.7 | 587 | 14.8 |
| **Total:** | **329,572** | **100.0** | **602** | **21.8** |

Table 10.A. Scale Score Distribution by Responses to Question 2 for Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Were any test questions different from the types of questions you see in your science class?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| **All** of the questions on the test were different. | 44,280 | 13.4 | 590 | 17.9 |
| **Most** of the questions on the test were different. | 129,401 | 39.3 | 603 | 20.7 |
| **Some** of the questions on the test were different. | 142,751 | 43.3 | 607 | 22.0 |
| **None** of the questions on the test were different. | 13,133 | 4.0 | 595 | 23.7 |
| **Total:** | **329,565** | **100.0** | **602** | **21.8** |

Table 10.A. Scale Score Distribution by Responses to Question 3 for Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **How hard were questions on this test compared to questions you see in your science class?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| They were **harder** than most questions in my science classes. | 136,453 | 41.4 | 596 | 18.8 |
| They were **about as hard** as the questions in my science classes. | 152,184 | 46.2 | 605 | 21.5 |
| They were **easier** than most questions in my science classes. | 40,926 | 12.4 | 614 | 25.3 |
| **Total:** | **329,563** | **100.0** | **602** | **21.8** |

Table 10.A. Scale Score Distribution by Responses to Question 4 for Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Do you think you will be enrolling in any more science classes in high school?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| Yes, I **will be** enrolling in additional high school science classes. | 155,656 | 47.2 | 608 | 22.5 |
| No, I will **not** be taking any additional high school science classes. | 173,906 | 52.8 | 597 | 19.5 |
| **Total:** | **329,562** | **100.0** | **602** | **21.8** |

Table 10.A. Scale Score Distribution by Responses to Question 1 for Grade Twelve

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **How many topics on the test were taught in your science classes?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| **All** of the topics on the test were taught in my science classes. | 10,505 | 10.7 | 598 | 25.8 |
| **Most** of the topics on the test were taught in my science classes, but not all of them. | 36,551 | 37.1 | 604 | 23.8 |
| **Some** of the topics on the test were taught in my science classes, but not most of them. | 44,268 | 44.9 | 597 | 18.7 |
| **None** of the topics on the test were taught in my science classes. | 7,254 | 7.4 | 585 | 14.2 |
| **Total:** | **98,578** | **100.0** | **599** | **21.9** |

Table 10.A. Scale Score Distribution by Responses to Question 2 for Grade Twelve

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Were any test questions different from the types of questions you see in your science class?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| **All** of the questions on the test were different. | 15,340 | 15.6 | 587 | 16.5 |
| **Most** of the questions on the test were different. | 35,852 | 36.4 | 599 | 20.5 |
| **Some** of the questions on the test were different. | 42,221 | 42.8 | 604 | 22.5 |
| **None** of the questions on the test were different. | 5,157 | 5.2 | 595 | 24.2 |
| **Total:** | **98,570** | **100.0** | **599** | **21.9** |

Table 10.A. Scale Score Distribution by Responses to Question 3 for Grade Twelve

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **How hard were questions on this test compared to questions you see in your science class?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| They were **harder** than most questions in my science classes. | 38,874 | 39.4 | 592 | 17.8 |
| They were **about as hard** as the questions in my science classes. | 45,799 | 46.5 | 601 | 21.4 |
| They were **easier** than most questions in my science classes. | 13,897 | 14.1 | 611 | 26.8 |
| **Total:** | **98,570** | **100.0** | **599** | **21.9** |

Table 10.A. Scale Score Distribution by Responses to Question 1 for High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **How many topics on the test were taught in your science classes?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| **All** of the topics on the test were taught in my science classes. | 40,675 | 9.0 | 600 | 25.5 |
| **Most** of the topics on the test were taught in my science classes, but not all of them. | 173,767 | 38.4 | 607 | 23.2 |
| **Some** of the topics on the test were taught in my science classes, but not most of them. | 210,893 | 46.6 | 599 | 19.1 |
| **None** of the topics on the test were taught in my science classes. | 27,442 | 6.1 | 587 | 14.6 |
| **Total:** | **452,777** | **100.0** | **601** | **21.8** |

Table 10.A. Scale Score Distribution by Responses to Question 2 for High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Were any test questions different from the types of questions you see in your science class?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| **All** of the questions on the test were different. | 62,804 | 13.9 | 589 | 17.6 |
| **Most** of the questions on the test were different. | 175,546 | 38.8 | 602 | 20.6 |
| **Some** of the questions on the test were different. | 195,311 | 43.1 | 606 | 22.1 |
| **None** of the questions on the test were different. | 19,101 | 4.2 | 595 | 23.7 |
| **Total:** | **452,762** | **100.0** | **601** | **21.8** |

Table 10.A. Scale Score Distribution by Responses to Question 3 for High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **How hard were questions on this test compared to questions you see in your science class?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| They were **harder** than most questions in my science classes. | 186,313 | 41.2 | 595 | 18.6 |
| They were **about as hard** as the questions in my science classes. | 209,419 | 46.3 | 604 | 21.5 |
| They were **easier** than most questions in my science classes. | 57,028 | 12.6 | 613 | 25.7 |
| **Total:** | **452,760** | **100.0** | **601** | **21.8** |

Table 10.A. Scale Score Distribution by Responses to Question 4 for Grades Ten and Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Do you think you will be enrolling in any more science classes in high school?** | **N** | **Percent** | **Scale Score Mean** | **Scale Score SD** |
| Yes, I **will be** enrolling in additional high school science classes. | 170,294 | 48.1 | 608 | 22.4 |
| No, I will **not** be taking any additional high school science classes. | 183,894 | 51.9 | 597 | 19.4 |
| **Total:** | **354,188** | **100.0** | **602** | **21.7** |

**Notes for table 10.A.22:**

* “N/A” is not applicable.
* The asterisk (\*) indicates that this is the correlation for grade ten and grade eleven students, since the fourth survey question did not apply to grade twelve students.
* Survey responses were coded as follows:

Question 1: 1 = None, 2 = Some, 3 = Most, and 4 = All.

Question 2: 1 = All, 2 = Most, 3 = Some, and 4 = None.

Question 3: 1 = Harder, 2 = About as hard, and 3 = Easier.

Question 4: 1 = Yes and 2 = No.

Table 10.A. Correlations Between Student Survey Responses and CAST Scale Scores

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Question 1** | **Question 2** | **Question 3** | **Question 4** |
| Grade 5 | −0.02 | 0.13 | 0.01 | N/A |
| Grade 8 | 0.17 | 0.17 | 0.07 | N/A |
| High school—Grade 10 | 0.08 | 0.11 | 0.20 | −0.29 |
| High school—Grade 11 | 0.16 | 0.18 | 0.27 | −0.27 |
| High school—Grade 12 | 0.16 | 0.21 | 0.29 | N/A |
| High school—All grades | 0.16 | 0.18 | 0.27 | \*−0.26 |

## Continuous and Systematic Improvement

The fifth operational administration of the California Science Test (CAST) occurred in 2023–‍24. Throughout the past six years, continuous efforts have been made to improve the assessments. This chapter summarizes accomplishments and ongoing improvements for CAST as well as strategies to implement possible future improvements.

### 2023–24 Feedback for Continuous Improvement Survey

The California Assessment of Student Performance and Progress (CAASPP) program annually solicits feedback from educators through the Feedback for Continuous Improvement Survey. Local educational agency (LEA) and test site staff, as well as test administrators and test examiners, were invited to participate in the 2023–24 Feedback for Continuous Improvement Survey. Its goal was to highlight successes and identify areas for improvement. A total of 3,162 survey respondents participated in this survey for the 2023–‍24 administration, compared to 3,869 respondents for the previous year. The California Department of Education (CDE) and ETS use key recommendations from educators to implement positive changes in the following administration year.

Educators provided valuable feedback for potential improvements to the future administration of the CAASPP and the English Language Proficiency Assessments for California (ELPAC) by reporting some lessons they learned in 2023–24. Based on those lessons and suggestions for improvement, the *CAASPP and ELPAC Feedback for Continuous Improvement Survey and Focus Groups Report* (CDE, 2024) presents recommendations for the CDE, with the goal of enhancing the administrative support provided to LEAs and schools for future CAASPP and ELPAC test administrations. Refer also to subsection [*5.3.4 Feedback for Continuous Improvement Survey*](#_Feedback_for_Continuous_1) for assessment-specific results.

### Communications

#### Updated Program Website

The CAASPP and ELPAC program websites will be combined into one website, the CAASPP & ELPAC Website. This site will feature responsive design, improved navigability, and a simplified layout.

### Reporting

End-of-year data files will be completed earlier to support the CDE goal of delivering statewide reporting earlier than in previous years.

### Test Design

ETS works in collaboration with the CDE in planning, proposing, evaluating, and improving the CAST design.

The operational test forms for the 2023–24 CAST administration were delivered in accordance with the approved blueprint, as were the test forms in previous administrations.

### Item Development Improvements

#### Item Content Specifications

The ETS content team collaborated with the CDE to revise the item content specifications, a set of 175 documents that make the alignment of all three dimensions of the California Next Generation Science Standards—disciplinary core ideas, science and engineering practices, and crosscutting concepts—clearer on CAST items. These item content specifications were first shared with the public during the 2019–20 test administration and were met with positive feedback.

#### Processes

Field test item data and additional information from prior CAST administrations have informed ETS and the CDE on improvements to development techniques that targeted features in all item types such as

* language load,
* language complexity, and
* directional language.

Both positive and negative exemplar items from prior CAST administrations were used in item developer training, including adjustments to how developers approach constructed-response items. Responses from students validated some aspects of development processes and pointed out deficits in other aspects of the assessments. One specific example of a process improvement was to provide more concise language in items to explain the item type functionality.

Performance tasks (PTs) were field-tested as parallel PTs, defined as two distinct Segment C PTs with identical stimuli but unique item sets. One goal of parallel PTs is to increase the likelihood of yielding an operational Segment B PT from among the pool of parallel PT items that aligns with the California State Board of Education–approved revised CAST blueprint and CDE guidelines.

Refer to subsection [*3.1.3 Incorporation into Item Development Processes*](#_Incorporation_into_Item) for more information on item specifications.

#### Item Bank

Work to refresh the CAST item bank will continue through subsequent development cycles with the goal of developing items of low, medium, and high complexity for the 2024–25 test administration.

Table 11.1 shows, for the last two item development cycles followed by full data analyses, the total number of unique items put on the forms, the number of rejected items, and the rejection rates. From prior test administrations to the 2023–24 cycle, the low rejection rates for both grade five and grade eight remained similar, while the high school grade band (grade ten, eleven, or twelve) showed a decrease.

Table 11.1 Item Development Results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Total Number of Unique Items Field-Tested in the 2022–23 Test Administration** | **Number of Rejected Items from 2022–23** | **Rejection Rate from 2023 Data Review** | **Total Number of Unique Items Field-Tested in the 2023–24 Test Administration** | **Number of Rejected Items from 2023–24 Cycle** | **Rejection Rate from 2024 Data Review** |
| Grade 5 | 36 | 1 | 3% | 96 | 2 | 2% |
| Grade 8 | 36 | 2 | 6% | 108 | 6 | 6% |
| High school | 96 | 3 | 3% | 60 | 3 | 5% |

### Psychometric Analyses

#### Updated Artificial Intelligence Scoring Criteria

In February 2024, discussion with the California Technical Advisory Group (CA TAG) resulted in the addition of the proportional reduction in mean squared error (PRMSE; Casabianca & McCaffrey, 2023; Casabianca et al., 2023) as a statistical metric to use for evaluation of artificial intelligence (AI) model performance. The PRMSE corrects for the reliability of the human ratings and so provides a clearer measure of AI score prediction accuracy. It can reveal evidence in support of AI scoring even when human ratings have lower reliability. The PRMSE statistic provides an estimate of the proportion of variance in human scores that can be accounted for by automated scores. In the context of kindergarten through grade twelve assessments, values of PRMSE of 0.70 or greater constitute satisfactory evidence for using the AI scores (ETS, 2021).

Another result of the CA TAG discussion in February 2024 was the clarification that student group |SMD| ≤ 0.10 is to be used as a threshold for flagging items for further investigation, not as an acceptance criterion.

#### Combined Item Analyses

Starting with the 2023–24 test administration, only the final item analysis (FIA) will be conducted for CAST, to shorten the duration of the item analyses using a streamlined and automated process.

In past test administrations, ETS conducted the preliminary item analysis for CAST before the data review meeting (DRM) and FIA after the DRM. Combining the analyses allows extra time to accumulate data, which is sufficient for the FIA to inform the DRM. One benefit of this approach is that it will improve the stability of the statistics, because more data will be used for the item analysis to inform the DRM. It also reduces the amount of preparation needed for DRM materials.

### Accessibility Resources

Like all CAASPP assessments, CAST is administered using the Test Delivery System (TDS) created by Cambium Assessment, Inc. As such, implementation of new computer-based universal tools, designated supports, and accommodations are aligned with the TDS.

The following changes will be implemented during the 2024–25 CAST administration:

* Text-to-speech will be added to the student sign-in workflow and be available for all students who select the [**Read Page**] button on the *Student Sign-In* and subsequent pages.
* The embedded American Sign Language accommodation will be enabled after the student selects an icon instead of a context menu item.
* Color contrast with non-embedded print-on-demand will no longer be listed as a non-embedded accommodation.
* The embedded text-to-speech accommodation will be enhanced in the following ways:
* Will include line tracking, which causes the entire line to be highlighted lightly and each spoken word to be highlighted in a contrasting color as it is read aloud
* Will be enabled after the student selects an icon instead of a context menu item
* The definition of the non-embedded alternate response options accommodation will be expanded to include augmentative and alternative communication devices.
* The definition of non-embedded speech-to-text will note that Spanish is available for CAST.
* The embedded and non-embedded word prediction accommodation will be renamed “word completion.”

### References

California Department of Education. (2024). *2023–24 CAASPP and ELPAC feedback for continuous improvement survey and focus groups report* [Unpublished manuscript]. Sacramento, CA: California Department of Education.

Casabianca, J. M., & McCaffrey, D. F. (2023, April). *On the proportional reduction in mean squared error in AI model evaluation*. Paper presented at the annual meeting of the National Council for Measurement in Education. Chicago, Illinois.

Casabianca, J. M., McCaffrey, D. F., Johnson, M., Ricker, K., Rotou, O., & Martineau, J. (2023). *Exploration of the proportional reduction in mean squared error for evaluating automated scores*. ETS Research Memorandum.

ETS. (2021). *Best practices for constructed-response scoring*. Princeton, NJ: ETS.

1. Data for 2023–24 was retrieved from the *CalEdFacts* web page on the CDE website. [↑](#footnote-ref-2)
2. This definition was retrieved from the CDE California Longitudinal Pupil Achievement Data System (CALPADS) web page on the CDE website. [↑](#footnote-ref-3)
3. These were the same as the calculators used during administration of the Smarter Balanced Summative Assessment for Mathematics. [↑](#footnote-ref-4)
4. The expandable items universal tool was turned on by the test administrator in the Test Administrator Interface. [↑](#footnote-ref-5)
5. These were the same as the mathematics tools used during administration of the Smarter Balanced Summative Assessment for Mathematics. [↑](#footnote-ref-6)
6. PDFs of the science charts are available for download from the Accessibility Resources web page on the CAASPP & ELPAC Website. [↑](#footnote-ref-7)
7. This technical report is based on the versions of the CDE Accessibility Matrix and the *Usability, Accessibility, and Accommodations Guidelines* that were available during the 2023–24 CAASPP administration. [↑](#footnote-ref-8)
8. An SVM performs classification by finding the hyperplane that maximizes the margin between two classes. The vectors (cases) that define the hyperplane are the support vectors (Vapnik, 1995). The Support Vector Regression is an extension of SVMs and uses the same principles as the SVM for classification, with only a few minor differences (Drucker et al., 1996). [↑](#footnote-ref-9)
9. Students who answered all items, defined as the minimum number of items across all forms, would not have their scores adjusted. Other students’ NR scores would be adjusted by the proportion of items they answered. [↑](#footnote-ref-10)
10. The timing data is based on capturing the amount of time spent on answering the item(s) on each page. [↑](#footnote-ref-11)