

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



# California Assessment of Student Performance and Progress California Science Test 2020–‍2021 Technical Report

**Submitted May 12, 2023**

**By ETS**



**Contract #CN150012**

Table of Contents

[Chapter 1: Introduction 1](#_Toc103172523)

[1.1. Background 1](#_Toc103172524)

[1.2. Test Purpose 2](#_Toc103172525)

[1.3. Test Structure 2](#_Toc103172526)

[1.4. Intended Population 3](#_Toc103172527)

[1.5. Intended Use and Purpose of Test Scores 3](#_Toc103172528)

[1.6. Testing Window 4](#_Toc103172529)

[1.7. Impact of the Novel Coronavirus Disease 2019 Pandemic 4](#_Toc103172530)

[1.7.1. Remote Testing Flexibility 4](#_Toc103172531)

[1.7.2. Test-Taking Rates in the 2020–2021 Administration 5](#_Toc103172532)

[1.8. Significant CAST Developments in 2020–2021 5](#_Toc103172533)

[1.8.1. Test Design 5](#_Toc103172534)

[1.8.2. Remote Testing 6](#_Toc103172535)

[1.8.3. Accessibility Resources 6](#_Toc103172536)

[1.9. Groups and Organizations Involved with the CAST 7](#_Toc103172537)

[1.9.1. California State Board of Education 7](#_Toc103172538)

[1.9.2. California Department of Education 7](#_Toc103172539)

[1.9.3. California Educators 7](#_Toc103172540)

[1.9.4. National Science Experts 7](#_Toc103172541)

[1.9.5. Contractors 8](#_Toc103172542)

[1.10. Systems Overview and Functionality 9](#_Toc103172543)

[1.10.1. Test Operations Management System 9](#_Toc103172544)

[1.10.2. Test Delivery System 9](#_Toc103172545)

[1.10.3. Practice and Training Tests 10](#_Toc103172546)

[1.10.4. California Educator Reporting System 10](#_Toc103172547)

[1.10.5. Constructed-Response Scoring Systems for ETS 10](#_Toc103172548)

[1.11. Overview of the Technical Report 12](#_Toc103172549)

[References 13](#_Toc103172550)

[Chapter 2: An Overview of CAST Processes 14](#_Toc103172551)

[2.1. Item Development 14](#_Toc103172552)

[2.1.1. Design Guidelines 14](#_Toc103172553)

[2.1.2. Content Guidelines 14](#_Toc103172554)

[2.1.3. Item Types Guidelines 14](#_Toc103172555)

[2.2. Test Assembly 15](#_Toc103172556)

[2.2.1. Test Blueprint 15](#_Toc103172557)

[2.3. Test Administration 15](#_Toc103172558)

[2.3.1. Test Security and Confidentiality 15](#_Toc103172559)

[2.3.2. Procedures to Maintain Standardization 16](#_Toc103172560)

[2.4. Fairness and Accessibility 17](#_Toc103172561)

[2.4.1. Universal Tools, Designated Supports, and Accommodations 17](#_Toc103172562)

[2.4.2. Description of Differential Item Functioning Analyses 19](#_Toc103172563)

[2.5. Scores 19](#_Toc103172564)

[2.5.1. Score Reporting 19](#_Toc103172565)

[2.5.2. Aggregation Procedures 19](#_Toc103172566)

[2.6. Overview of Psychometric Analyses 20](#_Toc103172567)

[2.6.1. Description of Classical Item Analyses 20](#_Toc103172568)

[2.6.2. Description of Item Response Theory Analyses 20](#_Toc103172569)

[References 22](#_Toc103172570)

[Appendix 2.A: Special Services Summaries 23](#_Toc103172571)

[Chapter 3: Item Development 27](#_Toc103172572)

[3.1. Use of Evidence-Centered Design 27](#_Toc103172573)

[3.1.1. Principles 27](#_Toc103172574)

[3.1.2. Theory of Action Model 27](#_Toc103172575)

[3.1.3. Incorporation into Item Development Processes 29](#_Toc103172576)

[3.1.4. Item Types and Features 30](#_Toc103172577)

[3.2. Guidelines 32](#_Toc103172578)

[3.2.1. Plan 32](#_Toc103172579)

[3.2.2. Process 32](#_Toc103172580)

[3.2.3. Specifications 33](#_Toc103172581)

[3.2.4. Selection of Item Writers 33](#_Toc103172582)

[3.2.5. Item Writer Training 34](#_Toc103172583)

[3.3. ETS Item Review Process 35](#_Toc103172584)

[3.3.1. Overview 35](#_Toc103172585)

[3.3.2. ETS Content Review 35](#_Toc103172586)

[3.3.3. ETS Accessibility Review 36](#_Toc103172587)

[3.3.4. ETS Editorial Review 36](#_Toc103172588)

[3.3.5. ETS Sensitivity and Fairness Review 36](#_Toc103172589)

[3.4. California Department of Education Review 36](#_Toc103172590)

[3.5. California Educator Review 37](#_Toc103172591)

[3.5.1. California Educators as Content Experts 37](#_Toc103172592)

[3.5.2. Composition of Item Review Panels 37](#_Toc103172593)

[3.5.3. Meetings for Review of CAST Field Test Items 38](#_Toc103172594)

[3.6. Data Review Meeting 39](#_Toc103172595)

[References 41](#_Toc103172596)

[Chapter 4: Test Assembly 42](#_Toc103172597)

[4.1. Test Design 42](#_Toc103172598)

[4.2. Test Blueprints and Other Content Specifications 43](#_Toc103172599)

[4.2.1. Test Blueprints 43](#_Toc103172600)

[4.2.2. Other Content Specifications 46](#_Toc103172601)

[4.3. Test Production Process 46](#_Toc103172602)

[4.3.1. Selection of Items 46](#_Toc103172603)

[4.3.2. Test Forms 47](#_Toc103172604)

[4.3.3. Psychometric Criteria and Identification of Eligible Items 47](#_Toc103172605)

[4.3.4. Content Review of Forms 50](#_Toc103172606)

[4.3.5. California Department of Education Review of Forms 50](#_Toc103172607)

[4.3.6. Configuration of the Test Delivery System 50](#_Toc103172608)

[4.3.7. Test Form Delivery 51](#_Toc103172609)

[4.4. Performance Expectation Coverage 51](#_Toc103172610)

[4.5. Special Version Forms 51](#_Toc103172611)

[4.5.1. Braille Form 51](#_Toc103172612)

[4.5.2. Emergency Paper–Pencil Form 52](#_Toc103172613)

[4.5.3. Forms with Accessibility Features Other Than Braille 52](#_Toc103172614)

[References 53](#_Toc103172615)

[Appendix 4.A: Performance Expectation Distribution for Segment A 54](#_Toc103172616)

[Blueprints 54](#_Toc103172617)

[Accessibility Information 60](#_Toc103172618)

[Chapter 5: Test Administration 73](#_Toc103172619)

[5.1. Overview 73](#_Toc103172620)

[5.1.1. Remote and In-Person Testing 73](#_Toc103172621)

[5.1.2. Student Test-Taking Requirements 73](#_Toc103172622)

[5.1.3. Test-Taking Summary 74](#_Toc103172623)

[5.1.4. Demographic Student Group Summaries 74](#_Toc103172624)

[5.2. Procedures to Maintain Standardization 76](#_Toc103172625)

[5.2.1. Local Educational Agency CAASPP Coordinator 76](#_Toc103172626)

[5.2.2. CAASPP Test Site Coordinator 76](#_Toc103172627)

[5.2.3. Test Administrators 77](#_Toc103172628)

[5.2.4. Instructions for Test Administrators 77](#_Toc103172629)

[5.3. Local Educational Agency Training 79](#_Toc103172630)

[5.3.1. Workshops, Virtual Training, and Webcasts 80](#_Toc103172631)

[5.3.2. Videos and Guides 80](#_Toc103172632)

[5.3.3. Training for Proper Identification and Assignment of Designated Supports and Accommodations 81](#_Toc103172633)

[5.3.4. Feedback for Continuous Improvement Survey 81](#_Toc103172634)

[5.4. Accessibility Resources 82](#_Toc103172635)

[5.4.1. Accessibility Resource Categories 83](#_Toc103172636)

[5.4.2. Identification 87](#_Toc103172637)

[5.4.3. Assignment 88](#_Toc103172638)

[5.4.4. Usage of Designated Supports and Accommodations 88](#_Toc103172639)

[5.5. Practice and Training Tests 90](#_Toc103172640)

[5.6. Test Security and Confidentiality 90](#_Toc103172641)

[5.6.1. ETS’ Office of Testing Integrity 90](#_Toc103172642)

[5.6.2. Procedures to Maintain Standardization of Test Security 91](#_Toc103172643)

[5.6.3. Test Security Monitoring 92](#_Toc103172644)

[5.6.4. Security of Electronic Files Using a Firewall 92](#_Toc103172645)

[5.6.5. Transfer of Scores via Secure Data Exchange 93](#_Toc103172646)

[5.6.6. Data Management in the Secure Database 93](#_Toc103172647)

[5.6.7. Statistical Analysis on Secure Servers 93](#_Toc103172648)

[5.6.8. Student Confidentiality 94](#_Toc103172649)

[5.6.9. Student Test Results 94](#_Toc103172650)

[5.6.10. Security and Test Administration Incident Reporting System Process 95](#_Toc103172651)

[5.6.11. Appeals 97](#_Toc103172652)

[References 99](#_Toc103172653)

[Appendix 5.A: Test-Taking Rates 100](#_Toc103172654)

[Appendix 5.B: Demographic Summary 106](#_Toc103172655)

[Chapter 6: Scoring and Reporting 119](#_Toc103172656)

[6.1. CAST Scoring Process 119](#_Toc103172657)

[6.1.1. Scoring for Constructed-Response Items 119](#_Toc103172658)

[6.1.2. Scoring for Selected-Response Items 132](#_Toc103172659)

[6.2. Student Test Scores 132](#_Toc103172660)

[6.2.1. Theta Scores 133](#_Toc103172661)

[6.2.2. Scale Scores for the Total Assessment 133](#_Toc103172662)

[6.2.3. Achievement Levels 135](#_Toc103172663)

[6.2.4. Domain Achievement Levels 136](#_Toc103172664)

[6.2.5. Theta Scores Standard Errors 140](#_Toc103172665)

[6.2.6. Scale Scores Standard Errors 141](#_Toc103172666)

[6.3. Reports Produced and Scores for Each Report 141](#_Toc103172667)

[6.3.1. Online Reporting 142](#_Toc103172668)

[6.3.2. Special Cases 142](#_Toc103172669)

[6.3.3. Types of Score Reports 143](#_Toc103172670)

[6.3.4. Score Report Applications 144](#_Toc103172671)

[6.3.5. Criteria for Interpreting Test Scores 144](#_Toc103172672)

[6.3.6. Criteria for Interpreting Score Reports 144](#_Toc103172673)

[6.4. New Artificial Intelligence Model Building 145](#_Toc103172674)

[6.4.1. Data Collection 145](#_Toc103172675)

[6.4.2. Model Training 145](#_Toc103172676)

[6.4.3. Model Evaluation 146](#_Toc103172677)

[References 147](#_Toc103172678)

[Accessibility Information 148](#_Toc103172679)

[Alternative Text for Equation 6.1 148](#_Toc103172680)

[Alternative Text for Equation 6.2 148](#_Toc103172681)

[Alternative Text for Equation 6.3 148](#_Toc103172682)

[Alternative Text for Equation 6.4 148](#_Toc103172683)

[Alternative Text for Equation 6.5 148](#_Toc103172684)

[Alternative Text for Equation 6.6 148](#_Toc103172685)

[Alternative Text for Equation 6.7 148](#_Toc103172686)

[Appendix 6.A: Overall Theta Score Distribution 149](#_Toc103172687)

[Appendix 6.B: Overall Scale Score Distribution 155](#_Toc103172688)

[Appendix 6.C: Demographic Summary of Overall Achievement Levels 173](#_Toc103172689)

[Appendix 6.D: Demographic Summary of Domain Achievement Levels 197](#_Toc103172690)

[Chapter 7: Psychometric Analyses 233](#_Toc103172691)

[7.1. Sample Used for Analyses 233](#_Toc103172692)

[7.2. Classical Item Analyses 233](#_Toc103172693)

[7.2.1. Classical Item Difficulty Indices (*p*-value) 234](#_Toc103172694)

[7.2.2. Item-Total Correlation 235](#_Toc103172695)

[7.2.3. Distribution of Item Scores 235](#_Toc103172696)

[7.2.4. Omit Rates 236](#_Toc103172697)

[7.2.5. Distractor Analyses 236](#_Toc103172698)

[7.2.6. Summary of Classical Item Analyses Flagging Criteria 236](#_Toc103172699)

[7.2.7. Classical Item Analyses Results 237](#_Toc103172700)

[7.3. Differential Item Functioning Analyses 238](#_Toc103172701)

[7.3.1. Differential Item Functioning Procedure for Dichotomous Items 238](#_Toc103172702)

[7.3.2. Differential Item Functioning Procedure for Polytomous Items 239](#_Toc103172703)

[7.3.3. Classification 240](#_Toc103172704)

[7.3.4. Differential Item Functioning Analysis Results 241](#_Toc103172705)

[7.4. Item Response Theory Analyses 242](#_Toc103172706)

[7.4.1. Models 243](#_Toc103172707)

[7.4.2. Data Preparation 244](#_Toc103172708)

[7.4.3. Calibration 244](#_Toc103172709)

[7.4.4. Equating 244](#_Toc103172710)

[7.4.5. Parameter Estimates 245](#_Toc103172711)

[7.5. Testing Time Analyses 246](#_Toc103172712)

[7.6. Reliability Analyses 250](#_Toc103172713)

[7.6.1. Test Score Reliability 250](#_Toc103172714)

[7.6.2. Constructed-Response Scoring Reliability 254](#_Toc103172715)

[7.7. Validity Evidence 256](#_Toc103172716)

[7.7.1. Evidence in the Design of the CAST 257](#_Toc103172717)

[7.7.2. Evidence Based on Test Content 258](#_Toc103172718)

[7.7.3. Evidence Based on Response Processes 259](#_Toc103172719)

[7.7.4. Evidence Based on Internal Structure 260](#_Toc103172720)

[7.7.5. Evidence Based on Relations to Smarter Balanced Test Scores 261](#_Toc103172721)

[7.7.6. Evidence Based on Consequences of Testing 262](#_Toc103172722)

[7.8. Score Comparability—Remote Versus In-Person Testing Analyses 262](#_Toc103172723)

[7.8.1. Analysis Sample 263](#_Toc103172724)

[7.8.2. Analyses Providing Context 264](#_Toc103172725)

[7.8.3. Student Performance on Individual Test Questions 265](#_Toc103172726)

[7.8.4. Student Performance on the Entire Test 265](#_Toc103172727)

[7.8.5. Summary of Reliability and Validity Analyses 266](#_Toc103172728)

[7.8.6. Limitations and Caveats for Data Interpretation 266](#_Toc103172729)

[References 267](#_Toc103172730)

[Accessibility Information 270](#_Toc103172731)

[Alternative Text for Equation 7.1 270](#_Toc103172732)

[Alternative Text for Equation 7.2 270](#_Toc103172733)

[Alternative Text for Equation 7.3 270](#_Toc103172734)

[Alternative Text for Equation 7.4 270](#_Toc103172735)

[Alternative Text for Equation 7.5 270](#_Toc103172736)

[Alternative Text for Equation 7.6 270](#_Toc103172737)

[Alternative Text for Equation 7.7 270](#_Toc103172738)

[Alternative Text for Equation 7.8 270](#_Toc103172739)

[Alternative Text for Equation 7.9 271](#_Toc103172740)

[Alternative Text for Equation 7.10 271](#_Toc103172741)

[Alternative Text for Equation 7.11 271](#_Toc103172742)

[Alternative Text for Equation 7.12 271](#_Toc103172743)

[Alternative Text for Equation 7.13 271](#_Toc103172744)

[Alternative Text for Equation 7.14 271](#_Toc103172745)

[Alternative Text for Equation 7.15 271](#_Toc103172746)

[Appendix 7.A: Item Difficulty Distribution 272](#_Toc103172747)

[Appendix 7.B: Item-Total Correlation Distribution 274](#_Toc103172748)

[Appendix 7.C: Item Discrimination Parameter Distribution 276](#_Toc103172749)

[Appendix 7.D: Item Difficulty Parameter Distribution 280](#_Toc103172750)

[Appendix 7.E: Testing Time Analyses 286](#_Toc103172751)

[Appendix 7.F: Reliability Analysis 291](#_Toc103172752)

[Appendix 7.G: Analysis of Classification 305](#_Toc103172753)

[Appendix 7.H: Correlations to Smarter Balanced Test Scores 309](#_Toc103172754)

[Chapter 8: Quality Control Procedures 321](#_Toc103172755)

[8.1. Quality Control of Test Materials 321](#_Toc103172756)

[8.1.1. Developing Computer-based Assessments 321](#_Toc103172757)

[8.1.2. Test Administration Instructions 321](#_Toc103172758)

[8.1.3. Collecting Test Materials 321](#_Toc103172759)

[8.1.4. Processing Test Materials 322](#_Toc103172760)

[8.2. Quality Control of Item Development 322](#_Toc103172761)

[8.3. Quality Control of Test Form Development 323](#_Toc103172762)

[8.4. Quality Control of Test Administration 323](#_Toc103172763)

[8.5. Quality Control of Scoring 324](#_Toc103172764)

[8.5.1. Quality Control of Machine-Scoring Procedures 324](#_Toc103172765)

[8.5.2. Quality Control of Human Scoring 324](#_Toc103172766)

[8.5.3. Artificial Intelligence Scoring Verification 325](#_Toc103172767)

[8.5.4. Enterprise Score Key Management System Processing 325](#_Toc103172768)

[8.6. Quality Control of Psychometric Processes 325](#_Toc103172769)

[8.6.1. Development of Scoring Specifications 325](#_Toc103172770)

[8.6.2. Psychometric Analyses 326](#_Toc103172771)

[8.7. Quality Control of Reporting 326](#_Toc103172772)

[8.7.1. Exclusion of Student Scores from Summary Reports 327](#_Toc103172773)

[8.8. Quality Control of End-to-End Testing 327](#_Toc103172774)

[8.8.1. Computer-based Assessments 327](#_Toc103172775)

[8.8.2. Paper–Pencil Tests 328](#_Toc103172776)

[Reference 329](#_Toc103172777)

[Chapter 9: Student Survey 330](#_Toc103172778)

[9.1. Student Survey Questions 330](#_Toc103172779)

[9.2. Student Survey Results 330](#_Toc103172780)

[Appendix 9.A: Results of Student Survey 332](#_Toc103172781)

[Chapter 10: Continuous and Systematic Improvement 342](#_Toc103172782)

[10.1. Test Design 342](#_Toc103172783)

[10.2. Item Development 342](#_Toc103172784)

[10.3. Administration and Test Delivery 343](#_Toc103172785)

[10.4. Constructed-Response Item Scoring 343](#_Toc103172786)

[10.4.1. Human Scoring Activities 343](#_Toc103172787)

[10.5. Psychometric Analyses 344](#_Toc103172788)

[10.6. Accessibility 344](#_Toc103172789)

[10.7. Feedback for Continuous Improvement Survey 345](#_Toc103172790)

[References 346](#_Toc103172791)

List of Tables

[Acronyms and Initialisms Used in the California Science Test Technical Report xiii](#_Toc103172792)

[Table 1.1 Number of Unique Items Assessed on the CAST 3](#_Toc103172793)

[Table 2.A.1 Special Services Summary for CAST, Grades Five and Eight—All Tested 23](#_Toc103172794)

[Table 2.A.2 Special Services Summary for CAST, High School Grades—All Tested 25](#_Toc103172795)

[Table 3.1 Item Types Used in the CAST 30](#_Toc103172796)

[Table 3.2 Total Number of Items Developed per Grade Level or Grade Band for the   
CAST 32](#_Toc103172797)

[Table 3.3 CAST Item Reviewer Qualifications 38](#_Toc103172798)

[Table 3.4 Item Data Review Results 40](#_Toc103172799)

[Table 4.1 CAST Blueprint for Segments Contributing to Individual Scores from the Physical Sciences Domain 44](#_Toc103172800)

[Table 4.2 CAST Blueprint for Segments Contributing to Individual Scores from the Life Sciences Domain 44](#_Toc103172801)

[Table 4.3 CAST Blueprint for Segments Contributing to Individual Scores from the Earth and Space Sciences Domain 45](#_Toc103172802)

[Table 4.4 CAST Blueprint for Segments Contributing to Individual Scores for All Grade Levels 45](#_Toc103172803)

[Table 4.5 The Number of Blocks for Each Segment for 2020–2021 CAST General Computer-based Forms 47](#_Toc103172804)

[Table 4.6 Desired Item Difficulty Distribution in Segment A 48](#_Toc103172805)

[Table 4.7 Theta Cut Scores Based on Standard Setting Results 49](#_Toc103172806)

[Table 4.8 Number of Items and Blocks Reviewed 49](#_Toc103172807)

[Table 4.9 PEs Assessed on the CAST 51](#_Toc103172808)

[Table 4.10 Cumulative Coverage of PEs 51](#_Toc103172809)

[Table 5.1 CAST Test-Taking Rates of the Full Population 74](#_Toc103172810)

[Table 5.2 Demographic Student Groups to Be Reported 75](#_Toc103172811)

[Table 5.3 Summary of Accommodations and Designated Supports Used by Students 89](#_Toc103172812)

[Table 5.4 Types of Appeals 96](#_Toc103172813)

[Table 5.5 Number of Appeals Requested in STAIRS for the 2020–2021 Administration—All Grade Levels 97](#_Toc103172814)

[Table 5.6 Number and Types of Incidents Submitted in STAIRS for the 2020–2021 Administration—All Grade Levels 98](#_Toc103172815)

[Table 5.A.1 CAST Test-Taking Rates for Grade Five by Student Group 100](#_Toc103172816)

[Table 5.A.2 CAST Test-Taking Rates for Grade Eight by Student Group 101](#_Toc103172817)

[Table 5.A.3 CAST Test-Taking Rates for Grade Ten by Student Group 102](#_Toc103172818)

[Table 5.A.4 CAST Test-Taking Rates for Grade Eleven by Student Group 103](#_Toc103172819)

[Table 5.A.5 CAST Test-Taking Rates for Grade Twelve by Student Group 104](#_Toc103172820)

[Table 5.A.6 CAST Test-Taking Rates for High School (All Grades Tested) by Student   
Group 105](#_Toc103172821)

[Table 5.B.1 Demographic Summary for Grade Five 106](#_Toc103172822)

[Table 5.B.2 Demographic Summary for Grade Eight 109](#_Toc103172823)

[Table 5.B.3 Demographic Summary for Grade Ten 111](#_Toc103172824)

[Table 5.B.4 Demographic Summary for Grade Eleven 113](#_Toc103172825)

[Table 5.B.5 Demographic Summary for Grade Twelve 115](#_Toc103172826)

[Table 5.B.6 Demographic Summary for High School (All Grades Tested) 117](#_Toc103172827)

[Table 6.1 CAST Sample Selection for Human-Scoring Procedures 121](#_Toc103172828)

[Table 6.2 Summary of Characteristics of Human Raters Scoring the CAST 123](#_Toc103172829)

[Table 6.3 Interrater Reliability and Descriptive Statistics for the Ratings by AI and Human Raters in AI-Scoring of Operational Items for Grade Five 127](#_Toc103172830)

[Table 6.4 Interrater Reliability and Descriptive Statistics for the Ratings by AI and Human Raters in AI-Scoring of Operational Items for Grade Eight 128](#_Toc103172831)

[Table 6.5 Interrater Reliability and Descriptive Statistics for the Ratings by AI and Human Raters in AI-Scoring of Operational Items for High School 129](#_Toc103172832)

[Table 6.6 Interrater Reliability and Descriptive Statistics for the Ratings by Two Raters in Human-Scoring of Operational Items for Grade Eight 130](#_Toc103172833)

[Table 6.7 Number of Operational CR Items Flagged by Scoring Method 131](#_Toc103172834)

[Table 6.8 Scaling Constants 133](#_Toc103172835)

[Table 6.9 Minimum Number of Item Requirements for Test Completion 134](#_Toc103172836)

[Table 6.10 Mean and SD of Theta Scores and Scale Scores 134](#_Toc103172837)

[Table 6.11 Scale Score Ranges for Achievement Levels 135](#_Toc103172838)

[Table 6.12 Percent of Students in Each Achievement Level for Total Scores 135](#_Toc103172839)

[Table 6.13 Description of Science Domain Achievement Levels 137](#_Toc103172840)

[Table 6.14 Percent of Students in Each Achievement Level for the Life Sciences   
Domain 139](#_Toc103172841)

[Table 6.15 Percent of Students in Each Achievement Level for the Physical Sciences Domain 139](#_Toc103172842)

[Table 6.16 Percent of Students in Each Achievement Level for the Earth and Space Sciences Domain 139](#_Toc103172843)

[Table 6.17 Number of Items for New AI Model Building by Grade Level 145](#_Toc103172844)

[Table 6.A.1 Overall Theta Score Distribution for Grade Five 149](#_Toc103172845)

[Table 6.A.2 Overall Theta Score Distribution for Grade Eight 150](#_Toc103172846)

[Table 6.A.3 Overall Theta Score Distribution for Grade Ten 151](#_Toc103172847)

[Table 6.A.4 Overall Theta Score Distribution for Grade Eleven 152](#_Toc103172848)

[Table 6.A.5 Overall Theta Score Distribution for Grade Twelve 153](#_Toc103172849)

[Table 6.A.6 Overall Theta Score Distribution for High School 154](#_Toc103172850)

[Table 6.B.1 Overall Scale Score Distribution for Grade Five 155](#_Toc103172851)

[Table 6.B.2 Overall Scale Score Distribution for Grade Eight 158](#_Toc103172852)

[Table 6.B.3 Overall Scale Score Distribution for Grade Ten 161](#_Toc103172853)

[Table 6.B.4 Overall Scale Score Distribution for Grade Eleven 164](#_Toc103172854)

[Table 6.B.5 Overall Scale Score Distribution for Grade Twelve 167](#_Toc103172855)

[Table 6.B.6 Overall Scale Score Distribution for High School 170](#_Toc103172856)

[Table 6.C.1 Percent of Students in Each Achievement Level for Total Scores by Demographic Variables for Grade Five 173](#_Toc103172857)

[Table 6.C.2 Percent of Students in Each Achievement Level for Total Scores by Demographic Variables for Grade Eight 177](#_Toc103172858)

[Table 6.C.3 Percent of Students in Each Achievement Level for Total Scores by Demographic Variables for Grade Ten 181](#_Toc103172859)

[Table 6.C.4 Percent of Students in Each Achievement Level for Total Scores by Demographic Variables for Grade Eleven 185](#_Toc103172860)

[Table 6.C.5 Percent of Students in Each Achievement Level for Total Scores by Demographic Variables for Grade Twelve 189](#_Toc103172861)

[Table 6.C.6 Percent of Students in Each Achievement Level for Total Scores by Demographic Variables for High School 193](#_Toc103172862)

[Table 6.D.1 Percent of Students in Each Achievement Level by Demographic Variables for Grade Five—Life Sciences Domain 197](#_Toc103172863)

[Table 6.D.2 Percent of Students in Each Achievement Level by Demographic Variables for Grade Eight—Life Sciences Domain 199](#_Toc103172864)

[Table 6.D.3 Percent of Students in Each Achievement Level by Demographic Variables for Grade Ten—Life Sciences Domain 201](#_Toc103172865)

[Table 6.D.4 Percent of Students in Each Achievement Level by Demographic Variables for Grade Eleven—Life Sciences Domain 203](#_Toc103172866)

[Table 6.D.5 Percent of Students in Each Achievement Level by Demographic Variables for Grade Twelve—Life Sciences Domain 205](#_Toc103172867)

[Table 6.D.6 Percent of Students in Each Achievement Level by Demographic Variables for High School—Life Sciences Domain 207](#_Toc103172868)

[Table 6.D.7 Percent of Students in Each Achievement Level by Demographic Variables for Grade Five—Physical Sciences Domain 209](#_Toc103172869)

[Table 6.D.8 Percent of Students in Each Achievement Level by Demographic Variables for Grade Eight—Physical Sciences Domain 211](#_Toc103172870)

[Table 6.D.9 Percent of Students in Each Achievement Level by Demographic Variables for Grade Ten—Physical Sciences Domain 213](#_Toc103172871)

[Table 6.D.10 Percent of Students in Each Achievement Level by Demographic Variables for Grade Eleven—Physical Sciences Domain 215](#_Toc103172872)

[Table 6.D.11 Percent of Students in Each Achievement Level by Demographic Variables for Grade Twelve—Physical Sciences Domain 217](#_Toc103172873)

[Table 6.D.12 Percent of Students in Each Achievement Level by Demographic Variables for High School—Physical Sciences Domain 219](#_Toc103172874)

[Table 6.D.13 Percent of Students in Each Achievement Level by Demographic Variables for Grade Five—Earth and Space Sciences Domain 221](#_Toc103172875)

[Table 6.D.14 Percent of Students in Each Achievement Level by Demographic Variables for Grade Eight—Earth and Space Sciences Domain 223](#_Toc103172876)

[Table 6.D.15 Percent of Students in Each Achievement Level by Demographic Variables for Grade Ten—Earth and Space Sciences Domain 225](#_Toc103172877)

[Table 6.D.16 Percent of Students in Each Achievement Level by Demographic Variables for Grade Eleven—Earth and Space Sciences Domain 227](#_Toc103172878)

[Table 6.D.17 Percent of Students in Each Achievement Level by Demographic Variables for Grade Twelve—Earth and Space Sciences Domain 229](#_Toc103172879)

[Table 6.D.18 Percent of Students in Each Achievement Level by Demographic Variables for High School—Earth and Space Sciences Domain 231](#_Toc103172880)

[Table 7.1 Item Difficulty Distributions 237](#_Toc103172881)

[Table 7.2 Item-Total Correlation Distributions 237](#_Toc103172882)

[Table 7.3 DIF Categories for Dichotomous Items 240](#_Toc103172883)

[Table 7.4 DIF Categories for Polytomous Items 240](#_Toc103172884)

[Table 7.5 Student Groups for DIF Comparison 240](#_Toc103172885)

[Table 7.6 Number of Items by DIF Category for Grade Five 241](#_Toc103172886)

[Table 7.7 Number of Items by DIF Category for Grade Eight 242](#_Toc103172887)

[Table 7.8 Number of Items by DIF Category for High School 242](#_Toc103172888)

[Table 7.9 Item Discrimination Parameter Distribution by Grade Level or Grade Band 245](#_Toc103172889)

[Table 7.10 Item Difficulty Parameter Distribution by Grade Level or Grade Band 246](#_Toc103172890)

[Table 7.11 Testing Time (in Minutes) for the Total Test 248](#_Toc103172891)

[Table 7.12 Summary Statistics for Scale Scores and Theta Scores, Reliability, and   
SEMs 251](#_Toc103172892)

[Table 7.13 Decision Accuracy for Reaching an Achievement Level 253](#_Toc103172893)

[Table 7.14 Decision Consistency for Reaching an Achievement Level 253](#_Toc103172894)

[Table 7.15 Frequencies of Ratings 254](#_Toc103172895)

[Table 7.16 Number of Students by Instructional Mode and Assessment Location 263](#_Toc103172896)

[Table 7.17 Average Item Parameter Estimates 265](#_Toc103172897)

[Table 7.18 Modeled Difference in Remote and In-Person Scale Scores: Remote Instruction Only 266](#_Toc103172898)

[Table 7.A.1 Item Difficulty Distributions by Item Type 272](#_Toc103172899)

[Table 7.A.2 Item Difficulty Distributions by Content Domain 273](#_Toc103172900)

[Table 7.B.1 Item-Total Correlation Distributions by Item Type 274](#_Toc103172901)

[Table 7.B.2 Item-Total Correlation Distributions by Content Domain 275](#_Toc103172902)

[Table 7.C.1 Item Discrimination Parameter Distribution by Item Type for Grade Five 276](#_Toc103172903)

[Table 7.C.2 Item Discrimination Parameter Distribution by Item Type for Grade Eight 277](#_Toc103172904)

[Table 7.C.3 Item Discrimination Parameter Distribution by Item Type for High School 277](#_Toc103172905)

[Table 7.C.4 Item Discrimination Parameter Distribution by Content Domain for Grade   
Five 278](#_Toc103172906)

[Table 7.C.5 Item Discrimination Parameter Distribution by Content Domain for Grade   
Eight 278](#_Toc103172907)

[Table 7.C.6 Item Discrimination Parameter Distribution by Content Domain for High   
School 279](#_Toc103172908)

[Table 7.D.1 Item Difficulty Parameter Distribution by Item Type for Grade Five 280](#_Toc103172909)

[Table 7.D.2 Item Difficulty Parameter Distribution by Item Type for Grade Eight 281](#_Toc103172910)

[Table 7.D.3 Item Difficulty Parameter Distribution by Item Type for High School 282](#_Toc103172911)

[Table 7.D.4 Item Difficulty Parameter Distribution by Content Domain for Grade Five 283](#_Toc103172912)

[Table 7.D.5 Item Difficulty Parameter Distribution by Content Domain for Grade Eight 284](#_Toc103172913)

[Table 7.D.6 Item Difficulty Parameter Distribution by Content Domain for High School 285](#_Toc103172914)

[Table 7.E.1 Testing Time (in Minutes) by Segment 286](#_Toc103172915)

[Table 7.E.2 Testing Time (in Minutes) by Item Type 289](#_Toc103172916)

[Table 7.F.1 Reliabilities and SEMs by Demographic Groups for Grade Five 291](#_Toc103172917)

[Table 7.F.2 Reliabilities and SEMs by Demographic Groups for Grade Eight 293](#_Toc103172918)

[Table 7.F.3 Reliabilities and SEMs by Demographic Groups for Grade Ten 295](#_Toc103172919)

[Table 7.F.4 Reliabilities and SEMs by Demographic Groups for Grade Eleven 297](#_Toc103172920)

[Table 7.F.5 Reliabilities and SEMs by Demographic Groups for Grade Twelve 299](#_Toc103172921)

[Table 7.F.6 Reliabilities and SEMs by Demographic Groups for High School 301](#_Toc103172922)

[Table 7.F.7 Reliabilities and SEMs by ELPAC Performance Levels 303](#_Toc103172923)

[Table 7.G.1 Grade Five: Decision Accuracy 305](#_Toc103172924)

[Table 7.G.2 Grade Five: Decision Consistency 305](#_Toc103172925)

[Table 7.G.3 Grade Eight: Decision Accuracy 305](#_Toc103172926)

[Table 7.G.4 Grade Eight: Decision Consistency 306](#_Toc103172927)

[Table 7.G.5 High School—Grade Ten: Decision Accuracy 306](#_Toc103172928)

[Table 7.G.6 High School—Grade Ten: Decision Consistency 306](#_Toc103172929)

[Table 7.G.7 High School—Grade Eleven: Decision Accuracy 306](#_Toc103172930)

[Table 7.G.8 High School—Grade Eleven: Decision Consistency 307](#_Toc103172931)

[Table 7.G.9 High School—Grade Twelve: Decision Accuracy 307](#_Toc103172932)

[Table 7.G.10 High School—Grade Twelve: Decision Consistency 307](#_Toc103172933)

[Table 7.G.11 High School: Decision Accuracy 307](#_Toc103172934)

[Table 7.G.12 High School: Decision Consistency 308](#_Toc103172935)

[Table 7.H.1 Correlations with Smarter Balanced ELA Test Scores for Grade Five 309](#_Toc103172936)

[Table 7.H.2 Correlations with Smarter Balanced Mathematics Test Scores for Grade   
Five 311](#_Toc103172937)

[Table 7.H.3 Correlations with Smarter Balanced ELA Test Scores for Grade Eight 313](#_Toc103172938)

[Table 7.H.4 Correlations with Smarter Balanced Mathematics Test Scores for Grade   
Eight 315](#_Toc103172939)

[Table 7.H.5 Correlations with Smarter Balanced ELA Test Scores for Grade Eleven 317](#_Toc103172940)

[Table 7.H.6 Correlations with Smarter Balanced Mathematics Test Scores for Grade   
Eleven 319](#_Toc103172941)

[Table 9.A.1 Scale Score Distribution by Responses to Question 1 for Grade Five 332](#_Toc103172942)

[Table 9.A.2 Scale Score Distribution by Responses to Question 2 for Grade Five 332](#_Toc103172943)

[Table 9.A.3 Scale Score Distribution by Responses to Question 3 for Grade Five 333](#_Toc103172944)

[Table 9.A.4 Scale Score Distribution by Responses to Question 1 for Grade Eight 333](#_Toc103172945)

[Table 9.A.5 Scale Score Distribution by Responses to Question 2 for Grade Eight 334](#_Toc103172946)

[Table 9.A.6 Scale Score Distribution by Responses to Question 3 for Grade Eight 334](#_Toc103172947)

[Table 9.A.7 Scale Score Distribution by Responses to Question 1 for Grade Ten 335](#_Toc103172948)

[Table 9.A.8 Scale Score Distribution by Responses to Question 2 for Grade Ten 335](#_Toc103172949)

[Table 9.A.9 Scale Score Distribution by Responses to Question 3 for Grade Ten 336](#_Toc103172950)

[Table 9.A.10 Scale Score Distribution by Responses to Question 4 for Grade Ten 336](#_Toc103172951)

[Table 9.A.11 Scale Score Distribution by Responses to Question 1 for Grade Eleven 337](#_Toc103172952)

[Table 9.A.12 Scale Score Distribution by Responses to Question 2 for Grade Eleven 337](#_Toc103172953)

[Table 9.A.13 Scale Score Distribution by Responses to Question 3 for Grade Eleven 338](#_Toc103172954)

[Table 9.A.14 Scale Score Distribution by Responses to Question 4 for Grade Eleven 338](#_Toc103172955)

[Table 9.A.15 Scale Score Distribution by Responses to Question 1 for Grade Twelve 339](#_Toc103172956)

[Table 9.A.16 Scale Score Distribution by Responses to Question 2 for Grade Twelve 339](#_Toc103172957)

[Table 9.A.17 Scale Score Distribution by Responses to Question 3 for Grade Twelve 340](#_Toc103172958)

[Table 9.A.18 Scale Score Distribution by Responses to Question 1 for High School 340](#_Toc103172959)

[Table 9.A.19 Scale Score Distribution by Responses to Question 2 for High School 341](#_Toc103172960)

[Table 9.A.20 Scale Score Distribution by Responses to Question 3 for High School 341](#_Toc103172961)

[Table 9.A.21 Scale Score Distribution by Responses to Question 4 for Grades Ten and Eleven 341](#_Toc103172962)

[Table 10.1 Item Development Results 343](#_Toc103172963)

List of Figures

[Figure 4.A.1 PE distribution for Segment A of the CAST grade five assessment 54](#_Toc103172964)

[Figure 4.A.2 PE distribution for Segment A of the CAST grade eight assessment 56](#_Toc103172965)

[Figure 4.A.3 PE distribution for Segment A of the CAST high school assessment 58](#_Toc103172966)

[Figure 6.1 Percentage of achievement levels 136](#_Toc103172967)

[Figure 6.2 Model building and evaluation process 146](#_Toc103172968)

Acronyms and Initialisms Used in the California Science Test Technical Report

|  |  |
| --- | --- |
| Term | Definition |
| 2PL-IRT | two-parameter logistic item response theory |
| ADEL | adult English learner |
| AERA | American Educational Research Association |
| AI | artificial intelligence |
| AIS | average item score |
| ALTRD | Assessment and Learning Technology Research & Development |
| APA | American Psychological Association |
| ASL | American Sign Language |
| CA NGSS | California Next Generation Science Standards |
| 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 |
| COVID-19 | novel coronavirus disease 2019 |
| 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 |
| 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 |
| IFEP | initial fluent English proficient |
| IEP | individualized education program |
| IMS | Instructional Management Systems |

Table of Acronyms and Initialisms *(continuation one)*

|  |  |
| --- | --- |
| Term | Definition |
| IRT | item response theory |
| ISAAP Tool | Individual Student Assessment Accessibility Profile Tool |
| JAWS | Job Access With Speech |
| K | kindergarten |
| KSAs | knowledge, skills, and abilities |
| LEA | local educational agency |
| LOSS | lowest obtainable scale score |
| LS | Life Sciences |
| MC | multiple choice |
| MCMS | multiple choice, multiple select |
| MCSS | multiple choice, single select |
| MH | Mantel-Haenszel |
| 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 |
| PIA | preliminary item analysis |
| PIN | problem item notification |
| PPT | paper–pencil test |
| 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 |

Table of Acronyms and Initialisms *(continuation two)*

|  |  |
| --- | --- |
| Term | Definition |
| SSR | Student Score Report |
| STAIRS | Security and Test Administration Incident Reporting System |
| SVM | support vector machine |
| TDS | test delivery system |
| TEI | technology-enhanced item |
| TIPS | Technology and Information Processing Services |
| TIF | test information function |
| TOMS | Test Operations Management System |
| UAT | user acceptance testing |
| US ED | US Education Department |
| 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, purpose of the test, intended population, impact of the novel coronavirus disease 2019 (COVID-19) pandemic, and organizations and systems involved.

### 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 through the use of a variety of item types and assessment approaches. These tests provide the foundation for the state’s school accountability system.

California adopted the California Next Generation Science Standards (CA NGSS) in September 2013. The CAST is a computer-based assessment aligned with the CA NGSS. It was administered as a pilot for the first time during the 2016–‍2017 CAASPP administration, followed by a field test administration during the 2017–‍2018 CAASPP administration. The first operational CAST was administered during the 2018–2019 CAASPP administration. The assessment is administered to students in grades five and eight and once in high school (i.e., in grade ten, eleven, or twelve [as long as the student is not repeating grade twelve]).

During the 2020–2021 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 grades
* California Alternate Assessments (CAAs) for ELA and mathematics in grades three through eight and grade eleven for students with significant cognitive disabilities
* Science assessments in grades five and eight and high school (grades ten, eleven, or twelve), including the CAST and the CAA for Science
* The California Spanish Assessment, optional for eligible students in grades three through eight and high school and designed to measure a student’s Spanish competency in 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

The 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. 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.

### Test Purpose

The purpose of the CAST is to assess students with federally required science assessments in grades five and eight and once in high school (i.e., grade ten, eleven, or twelve). The CAST is designed to assess the three dimensions (i.e., 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 (e.g., animations of scientific phenomena, virtual engineering challenges, or simulated experiments).

The 2020–2021 administration was the second operational year of the assessment and the second year that Student Score Reports (SSRs) were generated (refer to [*Chapter 6: Scoring and Reporting*](#_Scoring_and_Reporting)). Because of the very limited CAST administration in 2019–2020, SSRs were available only for students who submitted a test, and that year was not counted as an operational administration.

### Test Structure

The 2020–2021 test, administered at each grade level or grade band, comprised 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 tests.

At each grade level or grade band, the test delivery system (TDS) assigned students to two different operational item blocks in Segment A, with each containing 16–17 discrete items. Each test also contained two different operational PTs in Segment B, with each task presenting five to six items. Finally, either one field test discrete item block (with 13 discrete items) or one field test PT (with six to eight 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*](#_Item_Development_1)).

A braille form was available for students with visual impairments (refer to section[*2.3 Test Administration*](#_Test_Administration)). The braille form was composed of two Segment A blocks totaling 32–‍34 items and two Segment B PTs of 12–13 items.

Table 1.1 lists the total number of unique items per segment across test forms. (Braille computer-based forms and PPTs did not include Segment C.)

Table 1.1 Number of Unique Items Assessed on the CAST

|  |  |  |  |
| --- | --- | --- | --- |
| Segment | Grade 5 | Grade 8 | High School |
| A (2 blocks) | 33 | 34 | 34 |
| B (5–6 PTs) | 36 | 35 | 28 |
| C (8–9 discrete blocks) | 104 | 110 | 112 |
| C (5–11 PTs) | 52 | 35 | 72 |

### Intended Population

The CAST was expected to be administered to approximately 1.4 million students in the general population. The intended population was 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 section [*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 high school (grade ten, eleven, or twelve). Analyses of the results of the CAA for Science are reported separately.

### Intended Use and Purpose of Test Scores

The results of tests 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 Next Generation Science 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.”

In other words, results for tests within the CAASPP System are used for two primary purposes:

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

The 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 typical testing window starts in the middle of January and ends in the middle of July. However, because of the COVID-19 pandemic and the impacts to schools, the state testing window started on January 12 and ended on July 30, 2021. The July 30, 2021, date was an approved extension to the legislated testing window and offered LEAs more flexibility in testing, allowing schools more time for students to test.

Like other CAASPP assessments, the CAST was untimed for students. A student could take the CAST within the LEA’s 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 test was approximately two hours for grades five and eight and approximately one hour for high school.

### Impact of the Novel Coronavirus Disease 2019 Pandemic

#### Remote Testing Flexibility

When the 2020–2021 school year began, LEAs were offering varying instructional options, with a substantial percentage offering only distance learning options. When the US Department of Education (US ED) notified states that they should not expect waivers of the 2020–2021 annual state assessments, the CDE began exploring options for delivering its annual summative assessments. The approach taken by the CDE was to “allow flexibility for LEAs to utilize multiple test administration options to best meet the needs of students in response to the local context and to ensure the safety and health of students and LEA staff” (CDE, 2021a, p. 4). That flexibility offered LEAs two options for testing students using the CAASPP Assessments:

1. Test in person, with both students and test administrators co-located in the same room at a school or other secure location and following physical distancing guidelines.
2. Test remotely, with students and test administrators located at different physical locations. The test administrator would monitor students’ progress throughout the test by using remote monitoring tools connected to the TDS.

Note that remote testing was not available for students taking alternate assessments.

In addition, the flexibility included a local assessment option for the Smarter Balanced Summative Assessments. If it was not viable for an LEA to administer CAASPP assessments in person or remotely, an LEA could use a locally administered assessment (CDE, 2021a). A local assessment was not available for the CAST because of the newness of the CA NGSS. Finally, in April 2021, the US ED granted a waiver to the CDE for specific accountability and school identification requirements for the 2020–2021 school year, including removing the 95 percent participation rate penalty for the academic indicator (CDE, 2021c).

For all CAASPP assessments that offered remote testing, a comprehensive study was conducted to explore the validity of test score interpretation. In particular, the study evaluated whether the test options used in the 2020–2021 administration impacted student testing performance and whether the test options used in the 2020–2021 administration impacted student testing experience that, in turn, impacted the test score interpretation. Refer to subsection [*5.1.1 Remote and In-Person Testing*](#_Remote_and_In-Person_1) and section [*7.8 Score Comparability—Remote Versus In-Person Testing Analyses*](#_Score_Comparability—Remote_Versus) for detailed information about the CAST study.

#### Test-Taking Rates in the 2020–2021 Administration

The impact of the COVID-19 pandemic on the school year, and whether students returned to in-person learning, varied greatly across the state. The COVID-19 pandemic also impacted how students were tested, as remote administration was introduced and widely used.

As shown in table 5.1, test results were submitted for between approximately 14 percent of grade five students to approximately 29 percent of grade ten students. The percentage of students who submitted tests is lower than what would be observed during administration in a typical year because the CDE received a waiver from the federal government for the accountability of the statewide standardized assessments for the 2020–2021 school year. Unlike the 2019–2020 administration, when testing was terminated in early March 2020, the 2020–2021 CAST administration did produce sufficient data to conduct psychometric analyses and present results in this technical report.

### Significant CAST Developments in 2020–2021

#### Test Design

The 2020–2021 CAST administration featured the inaugural field testing of 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 California State Board of Education (SBE)-approved revised CAST blueprint and CDE guidelines, to go into effect in 2021–2022.

#### Remote Testing

LEAs were heavily impacted by the school and business shutdowns as a result of the COVID-19 pandemic and the impacts carried over into the beginning of the 2020–2021 CAASPP test administration. The CDE and ETS worked together to develop a way to allow schools to administer the CAASPP assessments to students remotely that included the following updates:

* The Student Testing Interface was enhanced to enable students and test administrators to access the tests securely when the student could not be tested at the school or LEA in person.
* The secure browser was updated with the following new features:
* One-on-one chatting or voice or video calls with the test administrator
* A raise-hand feature to signal the test administrator for attention
* Ability to approve a request from the test administrator to share a screen
* The Test Administrator Interface was updated with the following new features:
* Selection of an option to indicate whether the test session was in person or remote (for tracking and analysis purposes)
* One-on-one chatting or voice or video calls with a student
* One-way broadcast of messages to the students in the test session
* Multiple monitoring options (gallery or list views of testing students)
* Ability to request the student to share a screen
* Remote testing instructions and scripts were developed for test administrators, providing remote logon instructions.
* Remote testing administration videos were created to show test administrators how to give a remote test; how to monitor a remote test session; and how to schedule, start, and stop a remote test session.
* A video was created for students and parents/guardians to introduce them to remote testing.

#### Accessibility Resources

The following accessibility resource–related updates were made:

* CAASPP Matrix One has been combined with the English Language Proficiency Assessments for California Matrix Four to create the California Assessment Accessibility Resources Matrix that serves both testing programs.
* An increased number of prefetched items for braille embossing were sent to the embosser prior to the student’s reaching the item when the auto emboss feature was enabled for tests presented in braille. This feature enhancement allowed for items in fixed-form tests to be printed or embossed prior to the student’s reaching the item in the test, thus speeding up the testing time for students with the braille accommodation.
* A break was considered both an embedded and a non-embedded universal tool.
* The multiplication table accessibility resource was expanded to a 12 × 12 table.
* The description of the scratch paper resource indicated that students could create graphic organizers.

### Groups and Organizations Involved with the CAST

#### California State Board of Education

The 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 also is the state educational agency responsible for overseeing California’s compliance with the Every Student Succeeds Act as well as the state’s Public School Accountability Act that measures the academic performance and progress of schools on a variety of academic metrics (CDE, 2021e).

#### California Department of Education

The CDE oversees California’s public-school system, which is responsible for the education of more than 6,000,000 children and young adults in more than 10,500 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, 2021d).

#### California Educators

A variety of California educators, including teachers and school administrators—who were selected based on their qualifications, experiences, demographics, and geographic locations—were invited to participate in the various aspects of the assessment process. This included defining the purpose and scope, test design, item development, standard setting, score reporting, 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 the CAST.

#### Contractors

##### Primary Testing Contractor—ETS

The CDE and the SBE contract with ETS to develop and administer the CAST and report CAST results. As the primary testing contractor, ETS has the overall responsibility of working with the CDE to implement and maintain an effective assessment system and coordinating 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
* Providing a tiered help desk support system for LEAs
* 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)
* Developing all CAST items
* Scoring CR items
* Supporting the California Educator Reporting System (CERS)
* Constructing, producing, and controlling the quality of CAST test forms and related test materials
* Processing student test assignments
* Producing and distributing score reports electronically
* Completing all psychometric procedures
* Developing a summary score reporting website that can be viewed by the public

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

ETS also monitors and manages the work of Cambium Assessment, Inc. (CAI), ETS’ subcontractor 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;
* 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

### Systems Overview and Functionality

#### Test Operations Management System

TOMS is the password-protected, web-based system that LEAs use to manage all aspects of CAASPP testing. TOMS serves various functions for the CAST, 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
* Viewing and downloading reports
* Providing a platform for authorized user access to secure materials such as student data and results and access to the CAASPP Security and Test Administration Incident Reporting System/Appeals process

TOMS receives student enrollment data, including 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 CAST assessments—such as LEA CAASPP coordinators, CAASPP test site coordinators, test administrators, and test examiners—are assigned varying levels of access to TOMS. For example, only an LEA CAASPP coordinator has permission to set up the LEA’s test administration window; a test administrator cannot download student reports. A description of user roles is explained more extensively in the *2020–2021 CAASPP Online Test Administration Manual* (CDE, 2021b).

#### 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 tests and monitor student testing;
* the Student Testing Interface, on which students take the test using the secure browser; and
* the secure browser, the online 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

The publicly available practice and training tests were provided to LEAs to prepare students for administration of the CAST. These tests, available for grades five and eight and high school, simulated the experience of the CAST computer-based assessments. The training tests provided students with an opportunity to engage with the CAST technology and help them to become familiar with the CAST question types; the practice tests were intended to simulate the operational testing experience. The practice and training tests aligned with CA NGSS performance expectations but did not produce scores. Students, teachers, and the public could access them using a web browser, although accessing them through the secure browser permitted students to take the tests using the text-to-speech embedded accommodation and to test assistive technology. Test administrators could access scoring guides that described related scoring considerations.

The purposes of the practice and training tests are to

* allow students and test administrators to quickly become familiar with the user interface and components of the TDS and the process of starting and completing a testing session,
* introduce students and test administrators to grade-specific items similar to those on the operational assessment, which included discrete items and PTs, and
* 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.

Details on practice and training tests are presented in section [*5.5 Practice and Training Tests*](#_Practice_and_Training).

#### California Educator Reporting System

CERS is the system used by LEAs to view preliminary student results from the CAASPP assessments. The primary purpose of CERS is to provide educators and administrators with access to timely test results data for individual students and groups of students.

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

#### 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 6: Scoring and Reporting*](#_Scoring_and_Reporting).

For the 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 [*Chapter 8: Quality Control Procedures*](#_Quality_Control_Procedures).

The CR items could also be rated by AI scoring engines (e.g., the *c-*‍*rater*™ system). The use of such engines often required models be built with reliable human-rating data. For the 2017–2018 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 2018–2019, 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 for use in the 2019–2020 administration and for future administrations. A CR sampling plan to support the new AI model building is provided in subsection [*6.1.1.1.2 Sampling Process for Field Test Constructed-Response Items*](#_Sampling_Process_for). However, the 2019–2020 administration did not result in enough data for additional model building because of the COVID-19 pandemic, so no new models were built. AI scoring models were built and implemented using the 2020–2021 administration data for future use.

The *c-rater*™ engine is ETS’ system for the automated scoring of content in text-based responses. The engine uses state-of-the-art, machine-learning technology to score items that elicit and measure knowledge about specific content. 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 questions in content areas, including social studies, science, ELA, and mathematics.

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 (e.g., 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 model building step 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 2021 and contains nine additional chapters as follows:

* [Chapter 2](#_Overview_of_CAST) presents an overview of processes involved in a CAST testing cycle. This includes item development, test assembly and administration, and generation of test scores. It also includes information about the assignment of designated supports and accommodations.
* [Chapter 3](#_Item_Development_1) discusses the detailed procedures of item development for the 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](#_Chapter_5:_Test) details the processes involved in the administration of the CAST. It also describes the procedures followed by ETS to maintain test security throughout the test administration process.
* [Chapter 6](#_Scoring_and_Reporting) summarizes the types of scores and score reports that are produced at the end of each administration of the CAST.
* [Chapter 7](#_Chapter_7:_Analyses) summarizes the statistical procedures and results for 2020–2021. These analyses include

test taking rates,

classical item analyses,

differential item functioning analyses,

item response theory analyses,

response time analyses

reliability analyses, and

test location analyses

* [Chapter 8](#_Quality_Control_Procedures) highlights the quality control processes used at various stages of development and administration of the CAST.
* [Chapter 9](#_Student_Survey) describes the development and administration of the survey questionnaires for students and the results of analyses of their responses.
* [Chapter 10](#_Continuous_and_Systematic) discusses the various procedures used to gather information to improve the 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. (2021a). 2020–21 *Spring summative assessment administration flexibility guidelines.* California Department of Education website.

California Department of Education. (2021b). *CAASPP online test administration manual.* Sacramento, CA: California Department of Education.

California Department of Education. (2021c). *COVID-19 accountability FAQs.* California Department of Education website.

California Department of Education. (2021d, December). *Organization.* California Department of Education website.

California Department of Education. (2021e, September). *State Board of Education responsibilities.* California Department of Education website.

## An 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 incorporated innovations and best practices from national science assessments. For the CAST, items and associated stimuli that featured simulations were developed to integrate the dimensions of the performance expectations (PEs) while maintaining grade-level appropriateness for test takers. California science teachers assisted in creating these items, and 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 how California educators were selected and the process used to develop new items, mostly in 2019. Some item development activities were scaled back for the 2020–2021 administration as an impact of the ongoing novel coronavirus 2019 pandemic, with only newly added field test items for Segment C undergoing review, as described in [*3.5.1 California Educators as Content Experts*](#_California_Educators_as).

#### Design Guidelines

ETS content specialists referred 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 developed and used in the 2020–‍2021 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 adhered to ETS’ foundational guidelines for quality item writing. These guidelines formed the basis for training item writers and for the rigorous review process that was implemented for every item. Additionally, item content specifications and the CA NGSS PEs were used to guide the writing of items for the CAST. Refer to section [*3.2 Guidelines*](#_Item_Development) for the guidelines of item writing, including the item content specifications.

ETS trained California science teachers to develop items for the CAST during an item writing workshop in January 2019 (refer to subsections [*3.2.4 Selection of Item Writers*](#_Selection_of_Item) and [*3.2.5 Item Writer Training*](#_Item_Writer_Training)). California science teachers were instructed to produce items that spanned a variety of science and engineering practices and science domains (i.e., Life Sciences; Physical Sciences; Earth and Space Sciences; and Engineering, Technology, and Applications of Science) to provide as wide an array of items as possible for the CAST forms construction.

#### Item Types Guidelines

The 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 some new TEI types that used simulations and animations. Refer to section [*3.2 Guidelines*](#_Item_Development) 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 the CAST.

Greater emphasis was made to fill the CAST item bank with items that have students explore phenomena using item types that best fit the construct. A key factor in determining the assignment of PEs to each item writer was the teaching experience and expertise that the item writer possessed. ETS also generated item sets—PTs—internally to measure more complex skills in a particular domain.

### Test Assembly

The 2020–2021 CAST design was based on the State Board of Education–approved, high-level test design for an operational assessment, which requires that all students in the tested grades participate in three segments of the test: Segment A, Segment B, and Segment C. The first two segments comprised 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. The CAST had three main subcontent areas or domains: Life Sciences, Physical Sciences, and Earth and Space Sciences. The blueprints of the assessment are shown in table 4.1 through table 4.4.

For the 2020–2021 assessment, each student took 8 to 12 discrete items from each domain that were worth 12 to 18 total points and two PTs assessing different domains, worth 12 to 14 total points.

### Test Administration

The CAST was administered online 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 the CAST can be found in [*Chapter 5: Test Administration*](#_Chapter_5:_Test).

#### Test Security and Confidentiality

All operational tests within the California Assessment of Student Performance and Progress (CAASPP) System are secure. For the CAST, every person with access to test materials maintained the security and confidentiality of the tests. ETS’ internal Code of Ethics requires that all test information, including tangible materials (such as test booklets, test questions, and test results), confidential files, processes, and activities were kept secure. To ensure security for all tests that ETS develops or handles, ETS maintains an Office of Testing Integrity (OTI). A detailed description of the OTI and its mission is presented in subsection [*5.6.1 ETS’ Office of Testing Integrity*](#_ETS’s_Office_of) in [*Chapter 5: Test Administration*](#_Chapter_5:_Test).

In the pursuit of enforcing secure practices, ETS strives to safeguard the various processes involved in a test development and administration cycle. 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](#_Procedures_to_Maintain_1)
* [Test security monitoring](#_Test_Security_Monitoring)
* [Security of electronic files using a firewall](#_Security_of_Electronic_1)
* [Transfer of scores via secure data exchange](#_Transfer_of_Scores)
* [Data management in the secure database](#_Data_Management_in)
* [Statistical analysis on secure servers](#_Statistical_Analysis_on)
* [Student confidentiality](#_Student_Confidentiality)
* [Student test results](#_Student_Test_Results)

#### Procedures to Maintain Standardization

ETS takes all necessary measures to ensure the standardization of CAST administration. The measures for standardization include, but are not limited to, the aspects described in these subsections.

##### Test Administrators

The CAST grade and grade-level assessments are 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 Procedures to Maintain Standardization*](#_Procedures_to_Maintain).

Staff at local educational agencies (LEAs) involved in CAST administration include LEA CAASPP coordinators, CAASPP test site coordinators, and test administrators. The responsibilities of each of the staff members are described in the *CAASPP Online Test Administration Manual* (California Department of Education [CDE], 2021c).

##### Test Directions

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 is a web-based manual that provides test administration procedures and guidelines for LEA CAASPP coordinators and CAASPP test site coordinators, as well as the script and *Directions for Administration* to be followed exactly by test administrators during a testing session (CDE, 2021c). (Refer to [*5.2.4.2 CAASPP Online Test Administration Manual*](#_CAASPP_Online_Test) in [chapter 5](#_Chapter_5:_Test) for more information.)
* ***Spring Administration Information for Educators—***This is a web-based manual that was developed in response to the need to test students remotely. It supplements the *CAASPP Online Test Administration Manual* (CDE, 2021d). (Refer to [*5.2.4.3 Spring Administration Information for Educators*](#_Spring_Administration_Information) in [chapter 5](#_Chapter_5:_Test) for more information.)
* ***CAASPP and English Language Proficiency Assessments for California (ELPAC) Test Operations Management System (TOMS) User Guide*—**This is a web-based manual that provides instructions for TOMS, allowing LEA staff, including LEA CAASPP coordinators and CAASPP test site coordinators, to perform a number of tasks including setting up test administrations, adding and managing users, assigning tests, and configuring computer-based student test settings (CDE, 2021b). (Refer to [*5.2.4.4 CAASPP and English Language Proficiency Assessments for California (ELPAC) Test Operations Management System User Guide*](#_CAASPP_and_English_1) in [chapter 5](#_Chapter_5:_Test) for more information.)

### Fairness and Accessibility

All students enrolled in grades five and eight and once in high school (grade ten, eleven, or twelve [as long as the student is not repeating grade twelve]) are required to participate in the CAST, except for students with the most significant cognitive disabilities who meet the criteria for the California Alternate Assessment (CAA) for Science based on alternate achievement standards (approximately 1 percent or less of the student population). The decision to assign a student to take an alternate assessment is made by the student’s individualized education program (IEP) team.

There are several procedures in place to ensure that the CAST is fair and accessible to all test takers. This section provides information on the available accessibility resources.

All public school students participate in the CAASPP System of assessments (which includes the CAST), and includes 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 ELs and students with disabilities.

#### Universal Tools, Designated Supports, and Accommodations

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, 2020).

**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 CAST assessments 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 IEP or Section 504 plan.

**Accommodations** must be permitted on CAST assessments 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 only on approval by the CDE.

[Appendix 2.A](#_Appendix_2.A:_Special) presents counts and percentages of students assigned designated supports, accommodations, and unlisted resources for the 2020–2021 CAST administration. Table 2.A.1 presents data for grade five and eight tests, and table 2.A.2 presents data for grades ten, eleven, and twelve individually and aggregated for high school. The tables in [appendix 2.A](#_Appendix_2.A:_Special) were created using student demographic data in version 1 of the production data file (“P1”) updated on June 4, 2021.

##### Selection

The full list of the universal tools, designated supports, and accommodations used in all CAASPP computer-based assessments, including the CAST, is documented in the Accessibility Matrix (CDE, 2020). 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 the 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 Accessibility Matrix when deciding how best to support the student’s test-taking experience.

The Smarter Balanced Assessment Consortium’s *Usability, Accessibility, and Accommodations Guidelines* (“*Guidelines*”) (Smarter Balanced, 2020) aids in the selection of universal tools, designated supports, and accommodations deemed necessary for individual students.[[3]](#footnote-4) The *Guidelines* apply to all students and promote an individualized approach to the implementation of assessment practices. The *Guidelines* are intended to provide recommendations regarding universal tools, designated supports, and accommodations. Another manual, the *Smarter Balanced Usability, Accessibility, and Accommodations Implementation Guide* (Smarter Balanced, 2014),provides suggestions for implementation of these resources.

##### 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 CAASPP test site coordinator, either through individual assignment in 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 Individual Student Assessment Accessibility Profile (ISAAP) Tool. The ISAAP Tool could be used by LEAs in conjunction with the Guidelines and the CAASPP and ELPAC Accessibility Guide for Online Testing (CDE, 2021a), 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 [*1.10 Systems Overview and Functionality*](#_Systems_Overview_and) in [*Chapter 1: Introduction*](#_Introduction) for more details regarding the TDS.

##### Delivery

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 the computer-based CAST. Examples of embedded resources include the braille language resource, color contrast, and text-to-speech.

Non-embedded resources are available, when provided by the LEA, for both computer-based and paper–pencil CAASPP assessments. These resources are not part of the technology platform for the computer-administered CAASPP tests. 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) for a detailed description of the accessibility resources available to students taking the CAST.

#### Description of Differential Item Functioning Analyses

Differential item functioning (DIF) analyses are conducted to detect differences in student performance by identifying items for which one group of students performs significantly better than another group (e.g., male vs. female or White vs. Black or African American) after matching students on overall ability. 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 review these DIF-flagged items to determine the potential sources and meanings of performance differences. Refer to section [*7.3 Differential Item Functioning Analyses*](#_Differential_Item_Functioning_1) for more information.

### Scores

Individual student scores were reported for the 2020–2021 CAST administration. Student performance on the reporting scale was designated into one of the four achievement levels described in subsection [*6.2.3 Achievement Levels*](#_Achievement_Levels).

For information regarding score specifications and score reports, refer to [*Chapter 6: 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 the CAST at the student, school, and LEA levels. CERS also provides an individual score report. Refer to subsection [*6.3.1 Online Reporting*](#_Online_Reporting) for details about TOMS and CERS and subsection [*6.3.3 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 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 aggregated scores were presented for all students or selected demographic student groups.

Aggregated scores were generated by combining student scores. They can be created by combining results at the state, LEA or direct funded charter school, or school level; combining for all students; or by combining results for students who represent selected demographic student groups.

Aggregated results by demographic variables are presented in [appendix 6.C](#_Appendix_6.C:_Demographic). In table 6.C.1 through table 6.C.6 in [appendix 6.C](#_Appendix_6.C:_Demographic), students are grouped by demographic groups, including gender, ethnicity, English language fluency, special education service 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 fewer than 11. Definitions for the demographic student groups included in these tables are provided in table 5.2.

### Overview of Psychometric Analyses

Psychometric analyses were conducted on the CAST data, including classical item analyses, DIF analyses, item response theory (IRT) calibration, response time analyses, and reliability analyses. These analyses are described fully in [*Chapter 7: Psychometric Analyses*](#_Chapter_7:_Analyses). A brief description of classical item analyses and IRT analyses are provided in the following subsections.

ETS investigated test score validity and, in particular, whether the test options used in the 2020–2021 administration impacted student testing performance and whether the test options used in the 2020–2021 administration impacted student testing experience that, in turn, impacted the test score interpretation. A summary of this investigation is presented in section [*7.8 Score Comparability—Remote Versus In-Person Testing Analyses*](#_Score_Comparability—Remote_Versus).

#### 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 test takers 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 [*7.2 Classical Item Analyses*](#_Classical_Item_Analyses)*.* A description the differential item functioning analyses procedure is provided in section [*7.3 Differential Item Functioning Analyses*](#_Differential_Item_Functioning_1)*.*

#### Description of Item Response Theory Analyses

A concurrent calibration was conducted to estimate parameters for all items on the CAST. As a result of the concurrent calibration, the item parameter estimates were placed on a common scale for test items from the same grade-level or grade band assessment. The CAST is preequated, which means that for each administration, parameters were estimated only for the field test items, as the operational items already have parameter estimates.

The CAST forms for grades five and eight and high school were assembled with common items between the test versions, for each tested grade level and for high school, which support the efficiency and accuracy of the concurrent calibrations. The two-parameter logistic IRT model was used to calibrate the dichotomous items (i.e., items worth one point) and the generalized partial credit model (Muraki, 1992) was used to calibrate the polytomous items (i.e., items worth more than one point). Detailed procedures for the calibration and scaling analyses are included in section [*7.4. Item Response Theory Analyses*](#_Item_Response_Theory).

### References

California Department of Education. (2020). *California assessment accessibility resources matrix*. California Department of Education website.

California Department of Education. (2021a). *CAASPP and ELPAC* *accessibility guide for online testing.* Sacramento, CA: California Department of Education.

California Department of Education. (2021b). *CAASPP and ELPAC Test Operations Management System user guide*. Sacramento, CA: California Department of Education.

California Department of Education. (2021c). *CAASPP online test administration manual, 2019–20 administration*. Sacramento, CA: California Department of Education.

California Department of Education. (2021d). *Spring administration information for educators.* 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.

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

Smarter Balanced Assessment Consortium. (2014). *Smarter Balanced Assessment Consortium: Usability, accessibility, and accommodations implementation guide.* Los Angeles, CA: Smarter Balanced Assessment Consortium and National Center on Educational Outcomes.

Smarter Balanced Assessment Consortium. (2020). *Smarter Balanced Assessment Consortium: Usability, accessibility, and accommodations guidelines*. Los Angeles, CA: Smarter Balanced Assessment Consortium.

### Appendix 2.A: Special Services Summaries

Table 2.A.1 Special Services Summary for CAST, 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 | 24 | 0.04 | 14 | 0.02 |
| Embedded Accommodation—Braille | 1 | 0.00 | 1 | 0.00 |
| Non-Embedded Accommodation—Abacus | 39 | 0.06 | 8 | 0.01 |
| Non-Embedded Accommodation—Alternate Response Options | 24 | 0.04 | 21 | 0.03 |
| Non-Embedded Accommodation—Print on Demand | 26 | 0.04 | 15 | 0.02 |
| Non-Embedded Accommodation—Speech-to-Text | 1,011 | 1.62 | 961 | 1.28 |
| Non-Embedded Accommodation—Word Prediction | 512 | 0.82 | 440 | 0.59 |
| Embedded Designated Support—Color Contrast | 139 | 0.22 | 167 | 0.22 |
| Embedded Designated Support—Masking | 1,047 | 1.68 | 1,527 | 2.04 |
| Embedded Designated Support—Mouse Pointer (Size and Color) | 241 | 0.39 | 627 | 0.84 |
| Embedded Designated Support—Permissive Mode | 459 | 0.74 | 395 | 0.53 |
| Embedded Designated Support—Print Size | 245 | 0.39 | 206 | 0.27 |
| Embedded Designated Support—Stacked Translations | 174 | 0.28 | 252 | 0.34 |
| Embedded Designated Support—Streamlined Test Interface | 801 | 1.29 | 931 | 1.24 |
| Embedded Designated Support—Text-to-Speech | 4,969 | 7.98 | 4,789 | 6.39 |
| Embedded Designated Support—Translation Glossary | 257 | 0.41 | 350 | 0.47 |
| Embedded Designated Support—Turn off Universal Tools | 1 | 0.00 | 1 | 0.00 |

Table 2.A.1 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Accessibility Resource | Grade 5 Number | Grade 5 Pct. of Total | Grade 8 Number | Grade 8 Pct. of Total |
| Non-Embedded Designated Support—100s Number Table | 1,198 | 1.92 | 488 | 0.65 |
| Non-Embedded Designated Support—Amplification | 69 | 0.11 | 102 | 0.14 |
| Non-Embedded Designated Support—Calculator | 1,040 | 1.67 | 2,557 | 3.41 |
| Non-Embedded Designated Support—Color Contrast | 36 | 0.06 | 55 | 0.07 |
| Non-Embedded Designated Support—Color Overlay | 38 | 0.06 | 35 | 0.05 |
| Non-Embedded Designated Support—Magnification | 216 | 0.35 | 172 | 0.23 |
| Non-Embedded Designated Support—Medical Device | 15 | 0.02 | 17 | 0.02 |
| Non-Embedded Designated Support—Multiplication Table | 2,215 | 3.56 | 1,866 | 2.49 |
| Non-Embedded Designated Support—Noise Buffers | 1,264 | 2.03 | 968 | 1.29 |
| Non-Embedded Designated Support—Read Aloud | 1,528 | 2.46 | 1,215 | 1.62 |
| Non-Embedded Designated Support—Read Aloud in Spanish | 37 | 0.06 | 55 | 0.07 |
| Non-Embedded Designated Support—Science Charts | 1,070 | 1.72 | 1,766 | 2.35 |
| Non-Embedded Designated Support—Scribe | 466 | 0.75 | 225 | 0.30 |
| Non-Embedded Designated Support—Separate Setting | 3,999 | 6.43 | 4,628 | 6.17 |
| Non-Embedded Designated Support—Simplified Test Directions | 3,103 | 4.99 | 3,210 | 4.28 |
| Non-Embedded Designated Support—Translated Test Directions | 86 | 0.14 | 101 | 0.13 |
| Other—Unlisted Resources | 0 | 0.00 | 1 | 0.00 |
| Other—Designated support or accommodation is in IEP | 5,146 | 8.27 | 6,300 | 8.40 |
| Other—Designated support or accommodation is in Section 504 plan | 265 | 0.43 | 333 | 0.44 |

Table 2.A.2 Special Services Summary for CAST, 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 | 0 | 0.00 | 11 | 0.03 | 8 | 0.01 | 19 | 0.02 |
| Embedded Accommodation—Braille | 1 | 0.04 | 3 | 0.01 | 0 | 0.00 | 4 | 0.00 |
| Non-Embedded Accommodation—Abacus | 0 | 0.00 | 2 | 0.00 | 5 | 0.01 | 7 | 0.01 |
| Non-Embedded Accommodation—Alternate Response Options | 1 | 0.04 | 4 | 0.01 | 10 | 0.02 | 15 | 0.01 |
| Non-Embedded Accommodation—Print on Demand | 0 | 0.00 | 10 | 0.02 | 9 | 0.01 | 19 | 0.02 |
| Non-Embedded Accommodation—Speech-to-Text | 9 | 0.40 | 236 | 0.54 | 143 | 0.23 | 388 | 0.36 |
| Non-Embedded Accommodation—Word Prediction | 2 | 0.09 | 68 | 0.16 | 67 | 0.11 | 137 | 0.13 |
| Embedded Designated Support—Color Contrast | 1 | 0.04 | 23 | 0.05 | 2,027 | 3.28 | 2,051 | 1.90 |
| Embedded Designated Support—Masking | 9 | 0.40 | 152 | 0.35 | 2,255 | 3.65 | 2,416 | 2.24 |
| Embedded Designated Support—Mouse Pointer (Size and Color) | 0 | 0.00 | 15 | 0.03 | 63 | 0.10 | 78 | 0.07 |
| Embedded Designated Support—Permissive Mode | 4 | 0.18 | 60 | 0.14 | 20 | 0.03 | 84 | 0.08 |
| Embedded Designated Support—Print Size | 1 | 0.04 | 44 | 0.10 | 36 | 0.06 | 81 | 0.08 |
| Embedded Designated Support—Stacked Translations | 0 | 0.00 | 113 | 0.26 | 164 | 0.27 | 277 | 0.26 |
| Embedded Designated Support—Streamlined Test Interface | 4 | 0.18 | 97 | 0.22 | 102 | 0.17 | 203 | 0.19 |
| Embedded Designated Support—Text-to-Speech | 74 | 3.26 | 916 | 2.09 | 3,760 | 6.09 | 4,750 | 4.41 |
| Embedded Designated Support—Translation Glossary | 0 | 0.00 | 52 | 0.12 | 2,428 | 3.93 | 2,480 | 2.30 |
| Embedded Designated Support—Turn off Universal Tools | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |

Table 2.A.2 *(continuation)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| Non-Embedded Designated Support—100s Number Table | 1 | 0.04 | 85 | 0.19 | 74 | 0.12 | 160 | 0.15 |
| Non-Embedded Designated Support—Amplification | 0 | 0.00 | 23 | 0.05 | 19 | 0.03 | 42 | 0.04 |
| Non-Embedded Designated Support—Calculator | 48 | 2.11 | 697 | 1.59 | 988 | 1.60 | 1,733 | 1.61 |
| Non-Embedded Designated Support—Color Contrast | 0 | 0.00 | 7 | 0.02 | 7 | 0.01 | 14 | 0.01 |
| Non-Embedded Designated Support—Color Overlay | 0 | 0.00 | 5 | 0.01 | 7 | 0.01 | 12 | 0.01 |
| Non-Embedded Designated Support—Magnification | 0 | 0.00 | 34 | 0.08 | 27 | 0.04 | 61 | 0.06 |
| Non-Embedded Designated Support—Medical Device | 1 | 0.04 | 7 | 0.02 | 4 | 0.01 | 12 | 0.01 |
| Non-Embedded Designated Support—Multiplication Table | 8 | 0.35 | 192 | 0.44 | 224 | 0.36 | 424 | 0.39 |
| Non-Embedded Designated Support—Noise Buffers | 13 | 0.57 | 182 | 0.42 | 181 | 0.29 | 376 | 0.35 |
| Non-Embedded Designated Support—Read Aloud | 4 | 0.18 | 284 | 0.65 | 363 | 0.59 | 651 | 0.60 |
| Non-Embedded Designated Support—Read Aloud in Spanish | 0 | 0.00 | 7 | 0.02 | 69 | 0.11 | 76 | 0.07 |
| Non-Embedded Designated Support—Science Charts | 18 | 0.79 | 2,300 | 5.25 | 3,465 | 5.61 | 5,783 | 5.37 |
| Non-Embedded Designated Support—Scribe | 2 | 0.09 | 29 | 0.07 | 36 | 0.06 | 67 | 0.06 |
| Non-Embedded Designated Support—Separate Setting | 76 | 3.35 | 1,157 | 2.64 | 1,613 | 2.61 | 2,846 | 2.64 |
| Non-Embedded Designated Support—Simplified Test Directions | 22 | 0.97 | 600 | 1.37 | 2,904 | 4.70 | 3,526 | 3.27 |
| Non-Embedded Designated Support—Translated Test Directions | 0 | 0.00 | 56 | 0.13 | 101 | 0.16 | 157 | 0.15 |
| Other—Unlisted Resources | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Other—Designated support or accommodation is in IEP | 84 | 3.70 | 1,727 | 3.94 | 2,556 | 4.14 | 4,367 | 4.05 |
| Other—Designated support or accommodation is in Section 504 plan | 21 | 0.93 | 173 | 0.40 | 208 | 0.34 | 402 | 0.37 |

## Item Development

This chapter discusses the detailed procedures of item development for the 2020–2021 California Science Test (CAST) administration.

### Use of Evidence-Centered Design

#### Principles

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 actually know and can do. 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. By employing ECD during the development process, ETS built the validity argument needed to support the operational use of the 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 the CAST, including item development, so that the 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), the CAST allows students to demonstrate a wide array of scientific knowledge and skills. In this way, the 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 external interested educators to improve item alignment with the standards and provide students with grade-level appropriate phenomena. ETS incorporates feedback from external interested 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 the CAST and support the claims enumerated in the CAST blueprint.

The logic model provides the sequence of how the 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

The 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 (i.e., 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.

##### 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 in PTs that give the opportunity to experience authentic science.

***Educators*** will

* participate in item writing workshops to improve understanding in the multidimensionality of the CA NGSS,
* access item content specifications to develop aligned classroom assessment, and
* access released practice and training tests to better understand the item types and range of science content assessed on the CAST.

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

##### Intermediate Outcomes

After administration, the results from the 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 the 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 the CAST, teachers and students will have had limited access to teaching and learning of the CA NGSS. Therefore, score users may initially misinterpret 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 [*7.7.6 Evidence Based on Consequences of Testing*](#_Evidence_Based_on), 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 Achieve Next Generation Science Standards (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 grade band, 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 Specifications*](#_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 the CAST include assessment targets, framed from focal KSAs, for each dimension of the PE.

Similarly, the definition of claims for the 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, 2019).

#### 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 (i.e., SEP or CCC). Wherever possible, a single item assessed all three dimensions. 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 (e.g., 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 2020–‍2021 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 the CAST. This includes item types ranging from traditional multiple-choice (MC) items and constructed-response (CR) items (i.e., extended text) to new technology-enhanced item (TEI) types (the remainder of the item types).

Table 3.1 Item Types Used in the CAST

|  |  |
| --- | --- |
| Feature | Description |
| **Multiple Choice, Single Select (MCSS)** | This item type generally consists of a stem and list of choices; the test taker can select only one choice (option) to respond. The options have radio buttons, but the test taker 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)** | This item type generally consists of a stimulus, stem, and a list of choices; the test taker can select one or more choices (options) to respond. The options use check boxes, but the test taker can select text or an image by clicking in the option space rather than exactly in the check boxes. 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. |
| **Constructed Response (CR)** | This item type consists of a stem to which the test taker must provide a typed response in a designated response box. |

Table 3.1 *(continuation one)*

|  |  |
| --- | --- |
| Feature | Description |
| **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 test taker responds to this item type by placing a check mark in two or more cells in a grid. Grid items generally should have 3 rows and 3 columns and should not have more than 4 rows and 4 columns or 16 total checkboxes. |
| **Inline Choice List, Single Select** | In this item type, the stem contains a single blank that the test taker must fill in by selecting a single choice from a drop-down list. |
| **Inline Choice List, Multiple Select** | The stem contains two or more blanks that the test taker must fill in by selecting choices from a drop-down list of choices. |
| **Interactive** | Unscored simulations require the test taker to select inputs that provide certain outputs to answer one or more questions.  Scored simulations require the test taker to select inputs that are scored. |
| **Zone, Single Select** | This item type has answer choices that are predefined “hotspots” or zones on an image. When the test taker clicks on a hotspot, the selection is shaded light blue with a blue outline. The test taker selects one zone to respond. *This item type is no longer under development.* |
| **Match, Single Select** | The test taker responds to this item type by dragging and dropping a single choice (source) into the appropriate location (target). There should be a maximum of four sources, with only one of those sources to be placed in the single target. |
| **Match, Multiple Select** | The test taker responds to this item type 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. |
| **Bar-Picturegraph, Multiple Select** | The test taker responds to this item type 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 the CAST was developed through a comprehensive development cycle and designed to conform to principles of item writing defined by ETS. 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 the CAST focused on developing items that integrated at least two of the three dimensions of the CA NGSS—DCIs, 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 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 grade band to accommodate the CAST.

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

|  |  |  |  |
| --- | --- | --- | --- |
| Item Type | Grade 5 | Grade 8 | High School |
| Standard discrete item types (non-CR) | 75 | 77 | 77 |
| Discrete CR | 8 | 9 | 9 |
| PT items (3–8 tasks per grade) | 48 | 36 | 96 |
| **Total:** | **131** | **122** | **182** |

The standard discrete item types from table 3.2 include traditional MC items and familiar TEI types (e.g., match, inline choice list).

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

#### Process

Each CAST item was developed through a comprehensive development cycle and designed to conform to principles of quality item writing as defined by ETS. Further, each item in the CAST item bank was developed to measure a specific PE through integration of at least two of the three dimensions of the CA NGSS (i.e., DCI, CCC, and SEP). In addition, guidelines for style and for fairness—including issues related to bias and sensitivity—helped item developers and reviewers maintain item consistency across the item bank.

Throughout the item writing process, ETS adhered to its foundational guidelines for quality item writing. According to these guidelines, item developers conformed to the following list of eight attributes for each item:

1. The question 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.

#### Specifications

ETS created item content specifications for the 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.

#### Selection of Item Writers

For the item writing that occurred in 2019, senior ETS content staff screened applications for CAST item writers, and ETS approved only those with strong content and teaching backgrounds for the item writing training program. ETS selected item writers after the training, but not all recipients of the training became an item writer.

Because some of the participants were current or former California educators, they were particularly knowledgeable about the standards assessed by the CA NGSS. All item writers shared the following qualifications:

* Possession of a bachelor’s degree in science or in the field of education with special focus on a particular scientific domain (An advanced degree in the relevant content was desirable.)
* Previous experience or training in writing items for standards-based assessments, including knowledge of the many considerations that are important when developing items for special student populations
* Previous experience or training in writing items in the grades and content areas covered by the CAST
* Familiarity with, and understanding of, the CA NGSS

#### 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 the CAST, ETS recruited and trained science educators with diverse science backgrounds, including California teachers, to enrich the range of ideas brought to the process and support effective teaching practices in science.

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 the CAST; and
3. leverage the experiences, perspectives, and expertise of the teachers in writing items for the CAST.

ETS held an item writer–training workshop in January 2019 in Sacramento, California, 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. The dimensions of the CA NGSS (i.e., DCI, CCC, and SEP) were analyzed and explored to focus on the three dimensions of the CA NGSS that items for the CAST were to emphasize. To achieve this three-dimensional quality and maintain validity, ETS explained how items should elicit evidence of student reasoning instead of rote recall of science content associated with the DCI. Finally, ETS shared with trainees best practices in item writing to provide clarity within the item and avoid bias or sensitivity concerns.

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 alignment with the dimensions of the PEs in question 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 the 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 before presentation to the CDE and item review meetings, which are described in more detail in section [*3.5 California Educator Review*](#_California_Educator_Review_1).

The ETS review process for the CAST included the following four tasks; 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

CAST items and stimuli underwent three rounds of content reviews by content-area assessment specialists with increasing levels of expertise; these rounds are called Round 1, Round 2, and Final Round. These assessment specialists verified thatthe items and stimuli complied with the approved item specifications and with ETS’ written guidelines for clarity, style, accuracy, and appropriateness for California students, as well as complied with the task models. Assessment specialists reviewed each item for the following characteristics:

* Relevance of each item to the purpose of the test
* Match of each item to the task model, including depth of knowledge
* Match of each item to the principles of quality item writing
* Match of each item to the identified standard or standards
* Difficulty of the item
* Accuracy of the content of the item
* Readability of the item or passage
* Grade-level appropriateness of the item
* Appropriateness of any illustrations, graphs, or figures

Each item was classified with the PE that it was intended to measure. The assessment specialists checked each item against its classification codes, 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 had the choice to accept the item and classification as written, suggest revisions, or recommend that the item be discarded. These steps occurred prior to the CDE’s review.

#### 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.

#### ETS Editorial Review

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

#### ETS Sensitivity and Fairness Review

ETS assessment specialists who are specially trained to identify and eliminate questions that contain content or wording that could be construed to be offensive to, or biased against, members of specific student groups (e.g., ethnicity, race, or gender) conducted the next level of review (ETS, 2014, 2016a). These trained staff members reviewed every item before the CDE and item review panelist reviews.

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

* Cultural diversity
* Diversity of background, cultural tradition, and viewpoints to be found in the test‑taking populations
* Changing roles and attitudes toward various groups
* Role of language in setting and changing attitudes toward various groups
* 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 English learner (EL) students

### 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 item review meetings with California educators.

### California Educator Review

#### California Educators as Content Experts

In a typical year, ETS holds meetings with California educators at the end of the item review process as the final content-expert review that items must undergo before being placed 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 the CAST.

However, because of the novel coronavirus disease 2019 (COVID-19) pandemic, the suspension of testing in 2019–2020, and the decision prior to the conclusion of the item development process to reuse test forms from 2019–2020 for the 2020–2021 administration, a formal educator review meeting was not held. As part of the form reuse decision, ETS evaluated the CAST item bank to determine whether it would meet the needs for implementing the revised CAST blueprint approved in January 2020. The result of this evaluation found that additional field test items were needed to meet the revised blueprint for the 2021–2022 CAST administration. From May 11 to May 14, 2020, CDE staff facilitated meetings with 14 California educators to review these additional field test items for the 2020–2021 CAST test form so the 2021–2022 CAST test form could be built in alignment with the revised blueprint.

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 grade appropriateness as well as gender, racial, ethnic, or socioeconomic bias.

#### Composition of Item Review Panels

The panelists for typical item review meetings for CAST items included current teachers, resource specialists, administrators, curricular experts, and other education professionals. Minimum qualifications to be invited to participate were

* 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. As previously noted, a formal educator review meeting was not held. Table 3.3 shows the types of criteria captured for individuals who participate in the CAST item review. The data for a typical data review process would include numbers in the *Total* column.

Table 3.3 CAST Item Reviewer Qualifications

|  |  |  |
| --- | --- | --- |
| Qualification Type | Qualification | Total |
| **Occupation** | Special Education Teacher | - |
| **Occupation** | Educational Specialist | - |
| **Occupation** | General Education Teacher | - |
| **Highest Degree Earned** | Bachelor’s Degree | - |
| **Highest Degree Earned** | Master’s Degree | - |
| **Highest Degree Earned** | Doctorate | - |
| **Kindergarten (K)–12 Teaching Credential** | Elementary Teaching (multiple subjects) | - |
| **K–12 Teaching Credential** | Secondary Teaching (single subject) | - |
| **K–12 Teaching Credential** | Special Education | - |
| **K–12 Teaching Credential** | Reading Specialist | - |
| **K–12 Teaching Credential** | EL (Crosscultural, Language and Academic Development; Bilingual, Crosscultural, Language and Academic Development) | - |
| **K–12 Teaching Credential** | Administrative | - |
| **K–12 Teaching Credential** | Other | - |

During a typical item review process, item reviewers are recruited through an online application process. Recommendations are solicited from LEAs and county offices of education as well as from CDE and SBE staff. ETS assessment directors review applications and confirm that an applicant’s qualifications meet the specified criteria. Applicants who meet the criteria have their information forwarded to CDE staff for further review and agreement before invitations to participate are distributed.

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

Repercussions from the COVID-19 pandemic precluded ETS from facilitating a traditional item review meeting to support the 2020–2021 administration. Instead, the CDE hosted a discussion with select teachers in the field to review possible field test items aligned with PEs that were underrepresented in the CAST item bank. The discussion provided the CDE with sufficient recommendations on what to add to the suspended 2019–2020 field test Segment C to create the 2020–2021 field test Segment C.

Typically, ETS content-area assessment specialists facilitate CAST item review meetings. Each meeting begins with a brief training session on how to review and make recommendations for revising items.

ETS provides training on the following topics:

* Overview of the purpose and scope of the 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?

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.

ETS conducted an introductory training to highlight any new issues and serve as a statistical refresher. Reviewers then made recommendations about which items should be included in the item bank for future form assembly. If an 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 item to confirm there were no concerns, from a content perspective, as it pertained to the item flag. ETS psychometricians provided training on the item statistics and responded to questions about the item statistics during the item discussion. The data review meeting 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 items. The feedback was referenced when working with the CDE to reconcile educator feedback and to make a final decision on whether or not to include each field test item in the operational pool.

Refer to table 3.4 for the results of the item data review conducted for the 2020–2021 administration. This table shows the number of flagged items accepted without edits and the number of items rejected outright. The rejection rates were 13 percent, 12 percent, and 5 percent for grade five, grade eight, and high school, respectively.

Table 3.4 Item Data Review Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grade Level | Accept As Is | Reject | Total Items | Rejection Rate |
| Grade 5 | 33 | 5 | 38 | 13% |
| Grade 8 | 57 | 8 | 65 | 12% |
| High school | 91 | 5 | 96 | 5% |

### 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. (2019). *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 2020–2021 California Science Test (CAST) administration.

### Test Design

The 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 grades participate in three segments of the test: Segment A, Segment B, and Segment C. Segments A and B contribute to individual student score reporting. Segment C is used for field testing purposes, where the items do not count toward students’ score reporting. The braille computer-based forms and paper–pencil tests did not include field test items because of the increase in testing time needed by the students using the braille resource and that the same paper–pencil test 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.

As described in subsection [*3.5.1 California Educators as Content Experts*](#_California_Educators_as), the 2019–2020 test forms were reused during the 2020–2021 administration. However, to meet the revised blueprint for the 2021–2022 administration, some changes were made to the number of field test items on the forms administered in 2020–2021.

In the 2020–2021 design, Segment A contained two blocks of discrete items, each with 16 to 17 items, for a total of 42 to 44 points equally distributed between the two blocks. All students 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 two 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 4.1 through table 4.3) performance expectation (PE), a science domain for the item was designated based on 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 based on the predominant domain in the storyline. Each student received two PTs from two different domains out of the three main content domains—Life Sciences (LS), Physical Sciences (PS), and Earth and Space Sciences (ESS) (table 4.4). The PTs were presented after the discrete blocks and were delivered randomly, with either one delivered first.

Segment C contained eight or nine discrete item blocks and up to 11 PT blocks at each grade level or the grade band. Each student was administered one block with 13 discrete items or one PT with six to eight items. Sixteen of the 24 field test PT blocks came from the parallel PTs described in [*1.8.1 Test Design*](#_Test_Design).

### Test Blueprints and Other Content Specifications

#### Test Blueprints

Table 4.1 through table 4.4 show the CAST blueprint approved by the California SBE in November 2017 that was used to build the 2020–2021 operational forms. For details on the test blueprint for the CAST, refer to the *California Science Test Blueprint* (California Department of Education [CDE], 2017).

In these tables, an asterisk (\*) indicates that, across the three science content domains, a student will receive two to four items 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, 2020).

Table 4.1 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:** | **8–12 items** | **8–12 items** | **8–12 items** | **4–6 items per PT** |

Table 4.2 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:** | **8–12 items** | **8–12 items** | **8–12 items** | **4–6 items per PT** |

Table 4.3 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:** | **8–12 items** | **8–12 items** | **8–12 items** | **4–6 items per PT** |

Table 4.4 CAST Blueprint for Segments Contributing to Individual Scores for All Grade Levels

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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 |
| **PS** | 8–12 items | 12–18 points | 0–1 PTs | 4–6 items | 6–7 points |
| **LS** | 8–12 items | 12–18 points | 0–1 PTs | 4–6 items | 6–7 points |
| **ESS** | 8–12 items | 12–18 points | 0–1 PTs | 4–6 items | 6–7 points |
| **Total:** | **32–34 items** | **42–44 points** | **2 PTs** | **8–12 items** | **12–14 points** |

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

Segment A is designed to assess a student’s mastery of a breadth of PEs of the California Next Generation Science Standards in the grade bands tested (grade five, grade eight, and high school—either grade ten, eleven, or twelve).

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 had 8 to 10 items in each of the three science domains: Physical Sciences, Life Sciences, and Earth and Space Sciences. Two to four items were in 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.

#### Other Content Specifications

The two Segment A blocks were built to be parallel in terms of other content and statistical specifications.

Segment A had an item type distribution of 26 to 50 percent multiple-choice items and 38 to 58 percent technology-enhanced items per form. Each Segment A block had two to three constructed-response (CR) items. Most PTs contained one CR item. Cognitive complexity as measured by depth of knowledge (DOK) had a distribution for Segment A where up to 6 percent of the items on the form had DOK 1, 32 to 44 percent measured at DOK 2, and 50 to 62 percent measured at DOK 3 or 4.

The sequence number for items used in the operational form was as close as possible to the sequence number when the items were field-tested.

### Test Production Process

#### Selection of Items

Operational segments A and B in 2020–2021 were identical to those assembled for the suspended 2019–2020 CAST administration. To further build the CAST item bank, additional field test items (both discrete and PT) were added to the existing 2019–2020 Segment C to create the 2020–2021 Segment C. 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, test developers select items from the eligible item pool, ensuring that, as a whole, test forms

* meet the coverage specifications of the test blueprints (subsection [*4.2.1 Test Blueprints*](#_Test_Blueprints)),
* meet item selection criteria developed by the ETS psychometrics team (section [*4.3.3 Psychometric Criteria and Identification of Eligible Items*](#_Psychometric_Criteria_and)),
* represent a wide variety of item types, and
* provide a wide variety of item context.

#### Test Forms

Table 4.5 provides the number of blocks for each grade level and the grade band, for each segment of the general computer-based form.

Table 4.5 The Number of Blocks for Each Segment for 2020–2021 CAST General Computer-based Forms

|  |  |  |  |
| --- | --- | --- | --- |
| Block Description | Grade 5 | Grade 8 | High School (# of unique forms) |
| Operational Discrete Blocks | 2 | 2 | 2 |
| Operational PTs | 6 | 6 | 5 |
| Field Test Discrete Blocks | 8 | 9 | 9 |
| Field Test PTs | 8 | 5 | 11 |

The number of blocks summarized in table 4.5 resulted in the following number of unique forms for the general computer-based administration:

* Twelve forms for grade five
* Twelve 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

* One hundred ninety-two forms for grade five,
* One hundred sixty-eight forms for grade eight, and
* One hundred sixty 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 2020–2021 administration. Each of these two forms include two discrete blocks in Segment A and two 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 [*4.3.3 Psychometric Criteria and Identification of Eligible Items*](#_Psychometric_Criteria_and). The variation in the number of forms across grade levels is partially due to the number of operational and field test PTs available for each grade level.

#### Psychometric Criteria and Identification of Eligible Items

The classical item statistics such as the *p*-value (item difficulty; refer to subsection [*7.2.1 Classical Item Difficulty Indices (p-value)*](#_Classical_Item_Difficulty) for more details on this statistic), item total polyserial correlation (item discrimination; refer to subsection [*7.2.2 Item-Total Correlation*](#_Item-Total_Score_Correlations) for more details on this statistic), as well as the item response theory (IRT) parameters (refer to section [*7.4 Item Response Theory Analyses*](#_Item_Response_Theory) for more details on the IRT parameters) obtained from administrations of the CAST prior to the 2019–2020 administration would have been used to inform the item selection for the operational forms used in the 2020–2021 administration; however, due to the decision to reuse the 2019–2020 operational forms for the 2020–2021 administration, no additional operational item selections were made.

The following guidelines are typically 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 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 0.2. Most items selected had polyserial correlations higher than 0.3.
* Items flagged for C-DIF were not used unless it was necessary for content coverage (refer to subsection [*7.3.3 Classification*](#_Classification) 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 had 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 have been 2.5 or less.

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

* The average *b*-value of the form should have been around zero. The desired item difficulty distribution for the two Segment A blocks is shown in table 4.6. The item 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 have been given a higher priority.

Table 4.6 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 had parallel test information functions. For each block, test information should have been maximized around three theta points where the cut scores were set. Standard setting following the 2018–2019 administration resulted in the theta cut scores shown in table 4.7.

Table 4.7 Theta Cut Scores Based on Standard Setting Results

|  |  |  |  |
| --- | --- | --- | --- |
| Grade Level | 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 have been comparable to those from the past administrations, especially at the ability intervals with high density.
* The sequence of items used in the operational form should have been as close as possible to the sequence in the block when they were field-tested. Whenever possible, the position difference at the block level should have been no more than five items.

ETS test 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 4.1 through table 4.4 and the psychometric criteria as described in this subsection.

The psychometric review is an iterative process. The psychometric 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.[[4]](#footnote-5) The number of items and blocks reviewed is summarized in table 4.8. 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 4.8 Number of Items and Blocks Reviewed

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grade Level | Number of Operational Discrete Blocks | Number of Operational PTs | Number of Field Test Discrete Blocks | Number of Field Test PTs | Total Number of Items |
| Grade 5 | 6 | 10 | 8 | 8 | 318 |
| Grade 8 | 6 | 10 | 9 | 5 | 309 |
| High school | 6 | 9 | 9 | 11 | 342 |

#### Content Review of Forms

After psychometric approval, the proposed assessment underwent two additional content reviews and one editorial review. The content reviewers were test developers who had not previously worked on the development of the test forms they were reviewing and typically worked on assessment programs other than the CAST. 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 PE
* Verification of coverage of the PEs
* Identification of any possible grammatical or production errors

#### California Department of Education Review of Forms

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 clueing between items or statistical issues. 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 test 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 test delivery system (TDS). Unlike other stages of the test production process, this stage must occur prior to every administration of the CAST, even in the case of a form reuse. Therefore, the configuration of the TDS was done prior to the 2020–2021 administration.

Each item underwent an extensive platform review on different operating systems, such as Windows, Linux, and iOS, to ensure that the item looked 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 test that would be administered to the students. The CDE had to approve the CAST UAT before the test could be released for administration to students.

#### Test Form Delivery

Students were administered the discrete blocks first, followed by the PTs. Within each type of block (i.e., discrete blocks or PTs), the blocks were delivered in a random order.

### Performance Expectation Coverage

The various blocks of items that comprise each segment of the CAST covered an extensive range of PEs; these PEs at the operational item–pool level are shown for the grade levels and the grade band in table 4.9.

Table 4.9 PEs Assessed on the CAST

|  |  |  |  |
| --- | --- | --- | --- |
| Grade Level | PEs Assessed | PEs Available | Percent of PEs Assessed |
| Grade 5 | 35 | 45 | 78% |
| Grade 8 | 40 | 59 | 68% |
| 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 percentages from year one (2018–2019) and year two (2020–2021) that are presented in table 4.10 do not sum to the cumulative percentage because some of the same PEs were assessed in both years. For example, on the grade eight CAST, 66 percent of all available PEs were assessed in year one and 68 percent were assessed in year two, with several of the same PEs assessed in both years.

Table 4.10 Cumulative Coverage of PEs

|  |  |  |  |
| --- | --- | --- | --- |
| Grade Level | Percent of PEs Assessed in Year Two | Percent of PEs Assessed in Year Two | Cumulative Percent of PEs Assessed in Years One and Two |
| Grade 5 | 76% | 78% | 98% |
| Grade 8 | 66% | 68% | 92% |
| High school | 49% | 56% | 72% |

### Special Version Forms

#### Braille Form

ETS designed a braille form for students with visual impairment. The same segment A and B pool of items and PTs that were used for embedded designated supports and accommodations were also used for the braille form.

The 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 two PTs from two different science content domains. These two 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 impairment. Adaptation 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. Adaptation did not change the item type. Twinning an item meant the item utilized a different item type while maintaining the same construct and storyline of the original item. Whether items were adapted or twinned, the item construct and overall cognitive complexity was maintained as closely as possible with the original parent item.

#### Emergency Paper–Pencil Form

The 2020–2021 CAST administration included a paper-based emergency 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 form used in 2020–2021 was the same as the 2019–2020 paper–pencil form and complied with item selection and forms construction criteria noted in prior sections. Standard, large-print, and braille paper–pencil forms were developed.

#### Forms with Accessibility Features Other Than Braille

A subset of the general form blocks was used to provide accessible content for those students who had one or more designated support or accommodation assigned, 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) for a list of designated supports and accommodations available during the 2020–2021 CAST administration.

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

### References

California Department of Education. (2017). *California Science Test blueprint.* California Department of Education website.

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

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

#### Blueprints

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.

Refer to the [*Alternative Text for Figure 4.A.1*](#_Alternative_Text_for_22) for a description of this spreadsheet image.

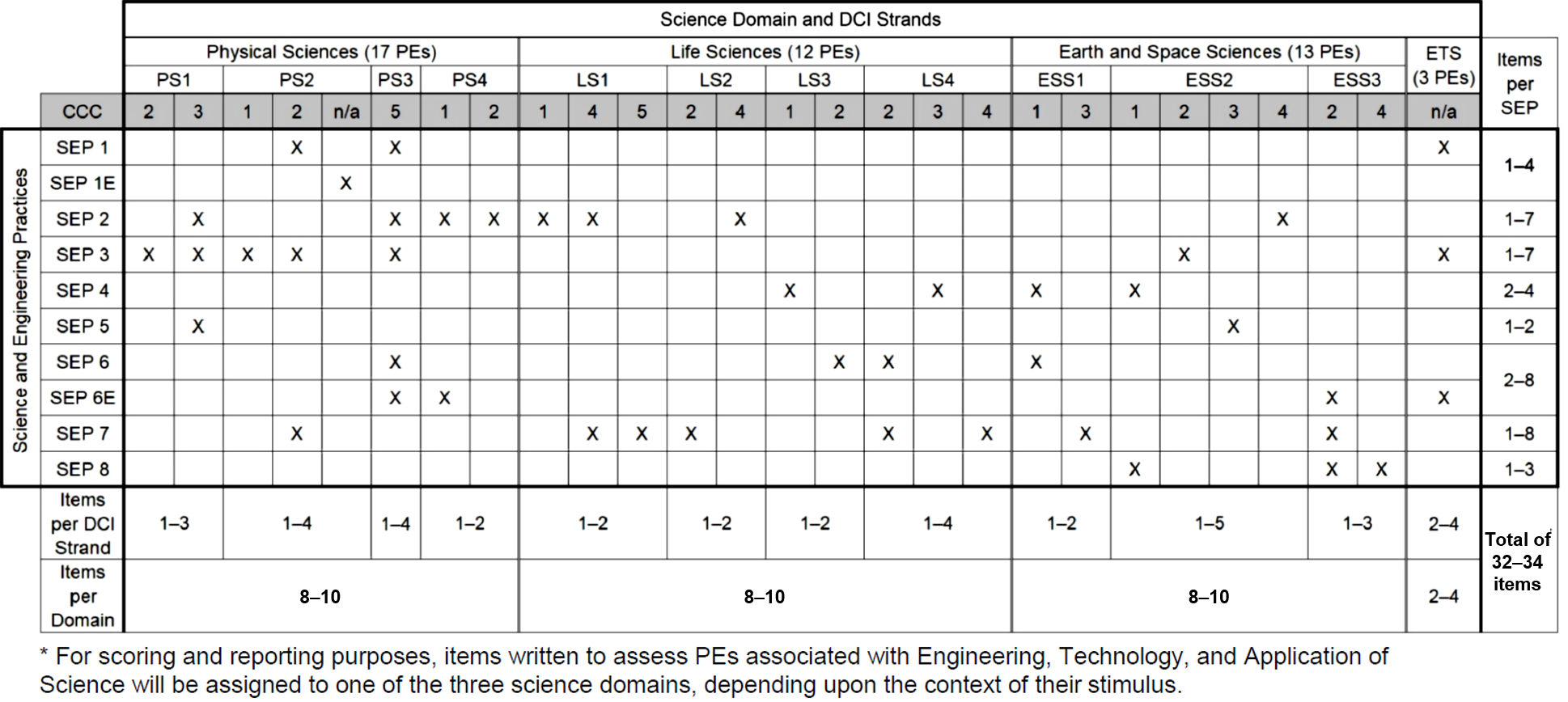


Figure 4.A.1 PE distribution for Segment A of the CAST grade five assessment

##### Notes on Figure 4.A.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.A.2*](#_Alternative_Text_for_23) for a description of this spreadsheet image.

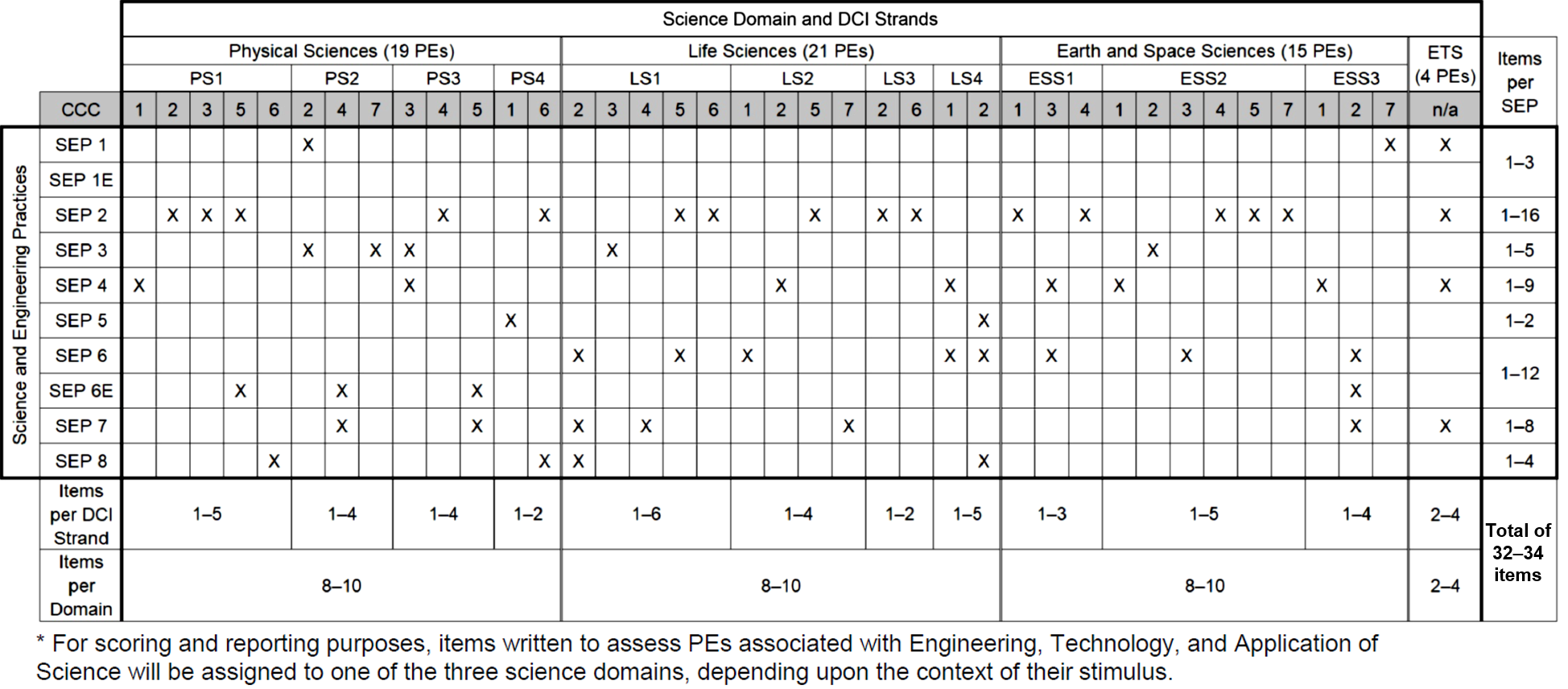


Figure 4.A.2 PE distribution for Segment A of the CAST grade eight assessment

##### Notes on Figure 4.A.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.A.3*](#_Alternative_Text_for_24) for a description of this spreadsheet image.

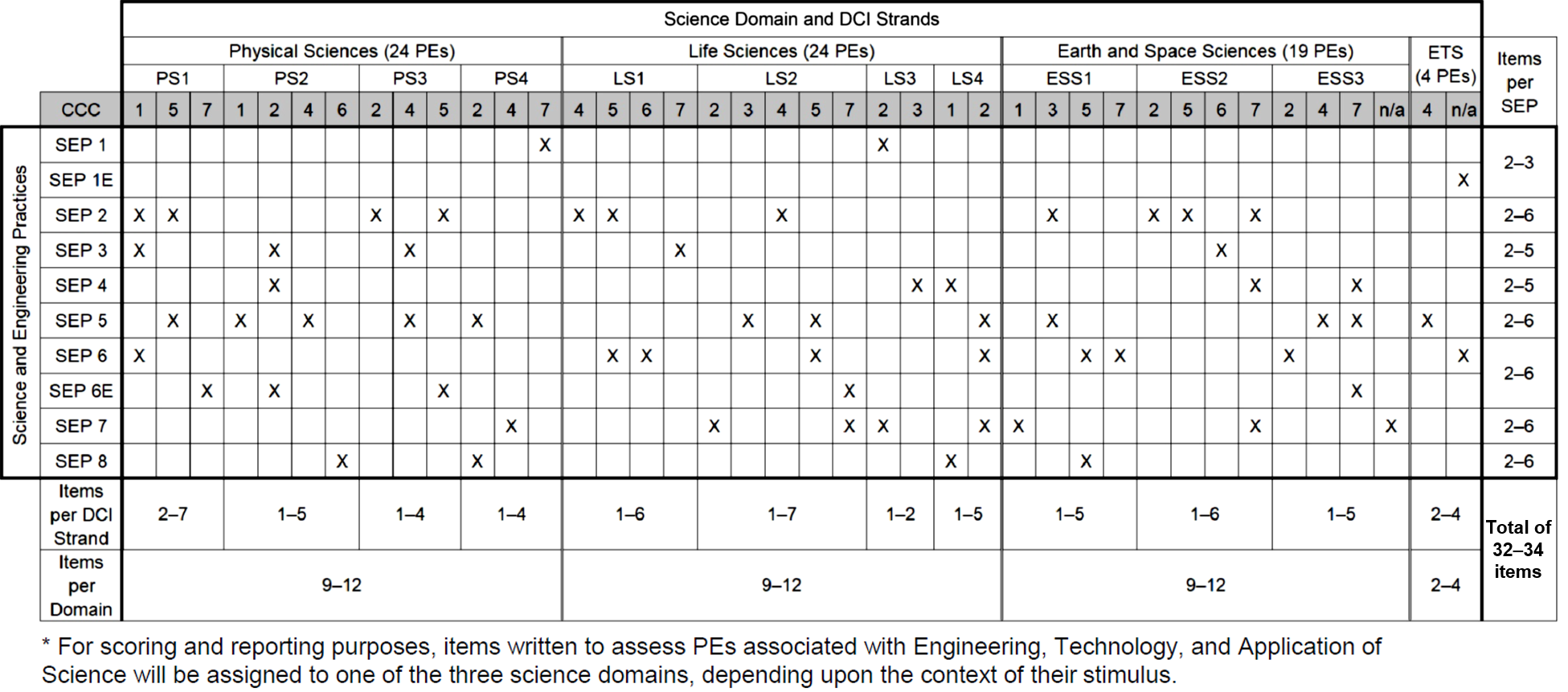


Figure 4.A.3 PE distribution for Segment A of the CAST high school assessment

##### Notes on Figure 4.A.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.

#### Accessibility Information

##### Alternative Text for Figure 4.A.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.A.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 8 and 10 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.A.1 *(continuation two)*

For the entire LS domain, between 8 and 10 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 8 and 10 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.A.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 32 to 34 items representing a selection of PEs across all three science domains and the ETS subdomain will be assessed on Segment A of the CAST.

##### Alternative Text for Figure 4.A.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.A.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.A.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.A.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 32 to 34 items representing a selection of PEs across all three science domains and the ETS subdomain will be assessed on Segment A of the CAST.

##### Alternative Text for Figure 4.A.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.A.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 8 and 10 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.A.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 8 and 10 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.A.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.A.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 to 34 items representing a selection of PEs across all three science domains and the ETS subdomain will be assessed on Segment A of the CAST.

## Test Administration

This chapter describes the details of California Science Test (CAST) administration, as well as the procedures followed by ETS to ensure test security.

### Overview

The CAST was administered to students in grades five and eight and in high school in spring 2021 in conjunction with the other tests that comprise the California Assessment of Student Performance and Progress (CAASPP) System.

In accordance with the procedures for all computer-based CAASPP assessments, local educational agencies (LEAs) identified test administrators to administer the CAST and entered them into the Test Operations Management System (TOMS). ETS provided LEA staff with the appropriate training materials, such as test administration manuals, videos, and webcasts, to ensure that the LEA staff and test administrators understood how to administer the computer-based CAST. The CAST used the same secure browser and computer-based testing platform as all the CAASPP assessments.

The testing window for the 2020–2021 administration of the CAST was planned for January 12 through July 30, 2021. Specific test administration schedules within that window were determined locally pursuant to the *California Code of Regulations*, Title 5 (5 *CCR),* sections 855(a)(1), 855(a)(2), 855(b), and 855(c).

#### Remote and In-Person Testing

When the 2020–2021 school year began, LEAs offered varying instructional options because of the novel coronavirus disease 2019 (COVID-19) pandemic, with a substantial percentage offering only distance learning. This resulted in the need for the California Department of Education (CDE) to explore different options for delivering its annual summative assessments. Two means of testing students were offered to LEAs for all CAASPP assessments (with the exception of the alternate assessments):

1. Test in person, with both students and test administrators co-located in the same room at a school or other secure location and following physical distancing guidelines
2. Test remotely, with students and test administrators located at different physical locations (The test administrator would monitor students’ progress throughout the test by using remote monitoring tools connected to the test delivery system [TDS])

After 2020–2021 testing, ETS conducted internal studies on the potential impact of the options provided for the 2020–2021 administration. The results of the analyses, when considering all students who tested, support the conclusion that remote testing can be viewed as reasonably comparable to in-person testing.

#### Student Test-Taking Requirements

The CAST was administered to students in grades five and eight as well as high school students in grade ten, eleven, or twelve who were assigned by their LEA. The CAST is a science assessment for the general student population (i.e., 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 assignment in 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 CDE suggesting that students who completed or were in the process of completing their last high school science course should take the science test, 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 test (CDE, 2019).

#### Test-Taking Summary

Table 5.1 presents the test-taking rates for all grade levels. Note that test takers are students who were registered for, and logged on to, the test. The percentage of registered students who submitted the CAST ranged from 13.75 percent for grade five to 28.51 percent for grade ten. The CDE received a waiver for the accountability of the statewide standardized assessments. Therefore, the percentage of students who tested is much lower than what would be expected during administration in a typical year.

Table 5.1 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 | 442,327 | 460,658 | 7,609 | 177,234 | 410,324 | 595,167 |
| Number of Started | 62,593 | 76,598 | 2,284 | 44,220 | 62,290 | 108,794 |
| Percent Started | 14.15 | 16.63 | 30.02 | 24.95 | 15.18 | 18.28 |
| Number of Expired | 1,751 | 7,161 | 115 | 2,419 | 3,157 | 5,691 |
| Percent Expired | 0.40 | 1.55 | 1.51 | 1.36 | 0.77 | 0.96 |
| Number of Submitted | 60,842 | 69,437 | 2,169 | 41,801 | 59,133 | 103,103 |
| Percent Submitted | 13.75 | 15.07 | 28.51 | 23.59 | 14.41 | 17.32 |

#### Demographic Student Group Summaries

Table 5.A.1 through table 5.A.6 in [appendix 5.A](#_Appendix_5.A:_Test-Taking) show the test-taking rates of selected demographic student groups for each test. The demographic student groups include economic status (disadvantaged or not), English language fluency, ethnicity, gender, homeless status, migrant status, parent/guardian military status, and special education services status.

Demographic student groups included in the summaries in this chapter are shown in table 5.2. The number and the percent of students for these demographic student groups are provided in [appendix 5.B](#_Appendix_5.B:_Demographic), starting in table 5.B.1 through table 5.B.5 for grade five and grade eight, and in table 5.B.6 for high school.

Table 5.2 Demographic Student Groups to Be Reported

|  |  |
| --- | --- |
| Category | Student Groups |
| **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 |
| **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** | * Military * Not military |
| **Special Education Services Status** | * Special education services * No special education services |

### Procedures to Maintain Standardization

The test administration procedures are designed so that the tests are administered in a standardized manner. ETS takes all necessary measures to ensure the standardization of test administration, as described in this section.[[5]](#footnote-6)

#### Local Educational Agency CAASPP Coordinator

An LEA CAASPP coordinator was designated by the district superintendent at the beginning of the 2020–2021 school year. LEAs include public school districts, 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 assessments. In addition to the responsibilities set forth in 5*CCR* Section 857, their responsibilities included

* adding CAASPP test site coordinators and test administrators into TOMS;
* training CAASPP test site coordinators and test administrators regarding the state requirements and CAASPP assessment administration, as well as security policies and procedures;
* reporting test security incidents (including testing irregularities) to the CDE using the online Security and Test Administration Incident Reporting System (STAIRS)/Appeals process;
* overseeing test administration activities;
* providing checklists for CAASPP test site coordinators and test administrators to review in preparation for administering the summative assessments;
* distributing and collecting scorable and nonscorable materials for students who take paper–pencil tests (PPTs); and
* requesting an Appeal (if indicated by TOMS prompts while reporting an incident using the STAIRS/Appeals process).

#### CAASPP Test Site Coordinator

A CAASPP test site coordinator is trained by the LEA CAASPP coordinator for each test site (5 *CCR* Section 857[f]). A test site 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]). CAASPP test site coordinators’ 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 all students take the test, 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 Administrators

Test administrators are identified by CAASPP test site coordinators as individuals who will administer the CAST.

A test administrator signed a security affidavit (5 *CCR* Section 850[ae]). 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 assessments, including the CAST;
* reporting all test security incidents to the test site 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 test with appropriate resources and reporting potential data errors to test site 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 the CAASPP (CDE, 2021e).

#### Instructions for Test Administrators

##### Test Administrator *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, 2021e). 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 TDS, so they could become familiar with the testing application used by their students (CDE, 2021e).

##### *CAASPP Online Test Administration Manual*

The *CAASPP Online Test Administration Manual* (CDE, 2021e) 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:

* Roles and responsibilities of those involved with CAASPP testing
* Test administration resources
* Test security
* Administration preparation and planning
* General 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 test and systems associated with the test.

##### *Spring Administration Information for Educators*

The *Spring Administration Information for Educators* (CDE, 2021f),which was developed in response to the need for remote administration of the CAASPP and English Language Proficiency Assessments for California (ELPAC), provided instructions and resources that coordinators and test administrators could use to prepare for testing and in test administration. Sections included the following topics:

* Administration options
* Requirements
* Test security
* Instructions for remote testing, including test administration directions and scripts
* Videos and quick reference guides
* Helpful links (including to the Parent/Guardian Information website)

##### *CAASPP and English Language Proficiency Assessments for California 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 PPTs.

TOMS modules described in the *TOMS User Guide* included the following (CDE, 2021d):

* **Test Administration Setup—**This module allowed LEAs to determine and calculate dates for the LEA’s 2020–2021 administration of the CAASPP, including the CAST.
* **Adding and Managing Users—**This module allowed LEA CAASPP coordinators to add CAASPP test site coordinators and test administrators to TOMS so that the designated user could administer, monitor, and manage the CAASPP, including the CAST.
* **Reports—**This module allowed LEA CAASPP coordinators and CAASPP test site coordinators access to the various reports in TOMS.
* **STAIRS/Appeals—**This module allowed LEA CAASPP coordinators and CAASPP test site coordinators access to create new STAIRS cases or search for STAIRS/Appeals cases.
* **Student Profile—**This module allowed LEA CAASPP coordinators, CAASPP test site coordinators, and test administrators and test examiners to view and manage students’ 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 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 assessments (CDE, 2021c).
* ***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, 2021b).
* ***CAASPP and ELPAC Accessibility Guide for Online Testing*—**This manual provided descriptions of the accessibility features for computer-based tests as well as information about supported hardware and software requirements for administering tests to students using accessibility resources, including those with a braille accommodation using Job Access With Speech (JAWS®) (software) or a braille embosser (hardware) (CDE, 2021a).

### 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 are developed with educator feedback and specify the audience, topics, frequency, and mode (in-person, virtual, videos, self-paced modules, etc.) of the training, including such elements as format, participants, and logistics.

In 2020–2021, ETS and SCOE quickly adapted training plans to meet the needs of educators deciding how to complete testing during the COVID-19 pandemic while adhering to local health guidance. All in-person trainings were converted to a virtual format, and the longer trainings were separated into shorter segments to avoid learner fatigue.

Knowing that educators were confronted with new challenges daily that put additional demands on their time, ETS and SCOE made every effort to make the information available in a variety of ways that allowed educators access to training at a time that was responsive to their varying circumstances. This included offering training events on multiple days and times, livestreaming events, recording and archiving trainings, and converting trainings to self-paced modules that could be taken any time, at the learner’s convenience.

All training opportunities were posted in one centralized location on the CAASPP website. LEA staff were able to register for training opportunities, across both CDE contractors’ offerings, in one place, on the Upcoming Training Opportunities web page. A Past Training Opportunities web page was also created, making it easier for educators to find missed training opportunities and providing easier access to recorded trainings.

#### Workshops, Virtual Training, and Webcasts

All offered virtual trainings were recorded and made available for on-demand viewing. Most trainings were offered via Zoom, a platform that educators quickly became familiar with and comfortable with during the COVID-19 pandemic. Zoom provided an opportunity for educators to ask questions and get answers in real time. Virtual trainings were also livestreamed on YouTube so that educators still had access if a particular training reached registration capacity.

In response to an environment where educators had competing priorities to juggle, ETS and SCOE employed a variety of strategies to increase engagement during virtual trainings. Live polls were 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 was enabled to give participants an opportunity to interact with each other or to provide open-ended feedback, or it was disabled to minimize distraction and drive attendees’ focus to the information being presented. Breakout groups were utilized in smaller group trainings, as appropriate. Breaks and processing time were incorporated into presentations to give attendees opportunities to attend to other responsibilities that might result as part of their job or home environment. Registered participants received an email from SCOE with a link to the virtual trainings.

Working closely with the CDE, ETS and SCOE were able to increase support to educators during a particularly challenging year. ETS offered weekly Office Hours and Coffee Sessions. Office Hours included CDE and ETS leadership to provide quickly changing updates on policies related to testing. Guest speakers from LEAs were invited to offer solutions and strategies for dealing with the challenges happening at the local level. Coffee Sessions included technical staff who could answer questions about all aspects of testing, including the newly offered remote testing option. SCOE continued to offer assessment update meetings intended to provide LEA coordinators with regular updates about California’s assessment system. All trainings and meetings were recorded and archived for on-demand viewing on the Past Training Opportunities web pages on the CAASPP website.

An unexpected benefit of the COVID-19 pandemic is that educators had greater access to CDE, ETS, and SCOE staff than they had in prior administration years. This challenging year provided an opportunity to provide more targeted support to educators that will have a lasting impact on the administrations to come.

#### Videos and Guides

To supplement the virtual trainings, ETS continued to produce videos on various aspects of administering the CAASPP and ELPAC assessments. 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 15 additional videos to support remote test administration. The videos included videos targeted to parents/‌guardians and students to provide instruction on how to download the secure browser on a personal device, so the assessment could be taken at home, and videos on how to take an assessment at home. Videos for parents/guardians and students were produced in both English and Spanish. SCOE produced a number of quick reference guides and guides to support remote testing, and those were made available in the 10 most common languages in California according to DataQuest.

#### Training for Proper Identification and Assignment of Designated Supports and Accommodations

ETS produced short demonstration videos for every embedded accessibility resource, demonstrating how to use the resource for educators, students, and parents/guardians. The videos were available in both English and Spanish on the Accessibility Resources Demonstration Videos web page on the CAASPP website. In addition, ETS also developed a video with LEA staff to help California educators learn more about the importance of implementing CAASPP accessibility resources and best practices used by educators in the field. The “Importance of Implementing CAASPP and ELPAC Accessibility Resources: Voices from Educators” video was available on the Quick Reference Guides and Videos web page on the CAASPP website.

Accessibility resource videos were also 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 2020–2021 CAASPP administration, ETS introduced a new virtual training series, “Matching Accessibility Resources to Students’ Needs.”This training focused on providing participants 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. The virtual training was originally intended as a one-time event but, because of overwhelming interest, the training was offered on four additional dates. The training was recorded and archived. LEA coordinators, test site coordinators, test administrators, and test examiners, were notified via email when the recorded training was available, further extending its reach.

At the California Assessment Conference, SCOE offered three sessions on accessibility. A “Plenary Accessibility 101” session was available as a prerecorded session for all conference attendees and was intended to build a shared understanding of basic accessibility-related terms and considerations. The “Digging Deeper into Accessibility” breakout session focused on developing an equitable and systematic process for matching students with appropriate accessibility resources. “Universal Design for Learning and Accessibility Resources: A Pathway to Success for All Students” was another breakout session focused on providing an opportunity to practice appropriately matching student needs to the various accessibility resources.

#### Feedback for Continuous Improvement Survey

ETS annually solicits feedback from educators through a survey that allows the CDE and ETS to focus on continuous improvement. LEA and test site staff, as well as test administrators and test examiners, were invited to participate in the 2020–2021 Feedback for Continuous Improvement Survey. Its goal was to highlight successes and identify areas for improvement.

Because of the unique nature of the 2020–2021 administration year and the option to administer assessments remotely or administer local assessments, the survey centered on preparation, training, and test administration, including remote testing. More than 1,600 California educators provided specific, actionable insights about their testing experience; in a more typical test administration year, 8,000 or more responses are generally received.

More than half (54%) of survey respondents used both remote and in-person options to complete testing. Overall, California educators continued to express positive experiences in their preparations for CAASPP and ELPAC administrations. Although the 2020–2021 administration included the daunting task of remote testing, educators felt that the resources and training materials they were given were useful in preparing them and their students for test administration. Their feedback generally described smooth preparation, training, support, and assessment administration experiences. Also, educators provided valuable feedback for potential improvements for future administrations based upon lessons learned.

The majority of respondents (64%) felt prepared for administering remote testing. On average, LEA CAASPP and ELPAC coordinators reported feeling prepared at a higher rate than those in site-level roles, such as test administrators and test examiners. Only a small percentage of respondents (6%) reported not feeling prepared. In regard to remote testing, educators felt they could benefit from more troubleshooting resources to deal with the technical difficulties that arise during remote testing. They indicated a simplified process for logging on to the secure browser would be helpful for students, particularly EL students. When asked about training preferences, respondents indicated that self-paced online trainings were preferable over in-person workshops and live virtual trainings. For live virtual trainings, Zoom is the preferred platform.

The CDE and ETS used key recommendations from educators to implement positive changes in the following administration year.

### Accessibility Resources

The U.S. Department of Education’s peer-review process includes several critical elements that address the need to monitor testing resources for students with disabilities, EL students, and EL students with disabilities. The Every Student Succeeds Act reaffirms the importance of ensuring that assessments are accessible to special 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 taking the CAST, as well as the procedures to identify and assign students with designated supports and accommodations. Finally, the number of students who were assigned accessibility resources was reported based on available data.

The 2020–2021 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 allow *all* students the opportunity to demonstrate what they know and what they are able to do, rather than giving students who use these resources an advantage over other students or artificially inflating their scores. 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’s California Assessment Accessibility Resources Matrix (Accessibility Matrix) (CDE, 2020a) 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 2020–2021 CAST administration.

###### Embedded

The following embedded universal tools were available to students testing in the secure browser:

* Breaks
* Calculator:[[6]](#footnote-7)

Four-function—grade five

Scientific—grade eight and high school

* Digital notepad
* English glossary
* Expandable items[[7]](#footnote-8)
* Expandable passages
* Highlighter
* Keyboard navigation
* Line reader
* Mark for review
* Mathematics tools (e.g., ruler, protractor)[[8]](#footnote-9)
* Science charts (i.e., calendar, Periodic Table of the Elements, conversion charts)
* Science tools (e.g., analog clock, laboratory equipment)
* Strikethrough
* Writing tools (e.g., 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 are accessibility features that were available for use by any student for whom the need had been indicated by an educator or a team of educators (with parent/‌guardian and student input as appropriate). Each designated support fell into one of two categories: embedded and non-embedded. Embedded designated supports were provided through the TDS (through the CAASPP secure browser).

The designated supports in the following subsections were available in the 2020–2021 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
* Stacked translations (Spanish)
* Streamline
* Text-to-speech (items and stimuli)
* Translations (glossary)
* 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 for items and stimuli
* Read aloud in Spanish
* Science charts (i.e., calendar, Periodic Table of the Elements, conversion charts)[[9]](#footnote-10)
* Scribe
* Separate setting (e.g., 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
* Magnification
* Masking
* Medical supports
* Multiplication table
* Noise buffers
* Read aloud (items)
* Scribe
* Separate setting (e.g., most beneficial time, special lighting or acoustics, adaptive furniture)
* Simplified test directions
* Translated test directions
* Translations (glossary)

##### Accommodations

Accommodations were changes in procedures or materials that increased equitable access during the CAASPP assessments. 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 2020–2021 CAST administration.

###### Embedded

The following embedded accommodations were available to students testing in the secure browser:

* American Sign Language (videos)
* Audio transcript (allowed but not currently used)
* Braille (embosser and refreshable)
* Closed-captioning (allowed but not currently used)

###### 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 resource that a student regularly uses in daily instruction, assessment, or both, that has not been previously identified as a universal tool, designated support, or accommodation. The Accessibility Matrix included an inventory of unlisted resources that were already identified and were preapproved (CDE, 2020). During the 2020–2021 CAST administration, an LEA CAASPP coordinator or CAASPP test site coordinator would use TOMS to submit a request for use of an unlisted resource. A preidentified, preapproved unlisted resource 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 measured were considered invalid for reporting purposes. The student’s score status would be changed to “Invalid” and the student’s scale score would be reported but appear on the Student Score Report (SSR) with an asterisk and a footnote that the test was administered under conditions that resulted in a score that may not be an accurate representation of the student’s achievement.

The CDE preidentified the following non-embedded unlisted resources that change the construct being measured:

* Bilingual dictionary
* English dictionary
* Signed exact English
* Thesaurus
* Translated word lists
* Translations

The LEA CAASPP coordinator or CAASPP test site 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. The lowest obtainable scale scores were reported for the affected domain when administrations included unlisted resources that changed the construct of that assessment.

#### Identification

All public school students participate in the CAASPP System, including students with disabilities and EL students. The Smarter Balanced Assessment Consortium’s *Usability, Accessibility, and Accommodations Guidelines* (Smarter Balanced, 2020) and the *California Assessment Accessibility Resources Matrix* (Accessibility Matrix) (CDE, 2020a) 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. The CAST follows the Smarter Balanced recommendations for use (Smarter Balanced, 2020).

The *Guidelines* apply to all students and promote an individualized approach to the implementation of assessment practices. Another web document, the *Smarter Balanced* *Resources and Practices Comparison Crosswalk* (Smarter Balanced, 2018),[[10]](#footnote-11) connects the assessment resources described in the *Guidelines* with associated classroom practices.

Another manual, the *Smarter Balanced Usability, Accessibility, and Accommodations Implementation Guide* (Smarter Balanced, 2014),provides suggestions for implementation of these resources. Test administrators are given the opportunity to administer the CAST practice and training tests so that students have the opportunity to familiarize themselves with a designated support or accommodation prior to testing.

#### Assignment

Once the student’s IEP or Section 504 plan team decided which accessibility resource(s) the student should use, LEA CAASPP coordinators and CAASPP test site coordinators used TOMS to assign designated supports and accommodations to students prior to the start of a test session.

There were three ways the 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 [*2.4.1.1 Selection*](#_Resources_for_Selection_1).)
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 CAASPP test site 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.

#### Usage of Designated Supports and Accommodations

LEA CAASPP coordinators and CAASPP test site 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 test 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 test so the student could be retested without the accommodation or designated support.

After schools and LEAs assigned eligible students to accommodations or designated supports, Cambium Assessment, Inc.’s (CAI’s) TDS provided and captured whether a certain accommodation or designated support, or multiple accommodations or designated supports, were used by a student as the student progressed through the test.

Table 5.3 reports the number of students who, based on the availability of data, were assigned to each accommodation or designated support and the number of students who actually used this accommodation or designated support at least once. 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 5.3 are as follows:

* **Text-To-Speech:** Text is read aloud to the student via embedded text-to-speech technology.
* **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.
* **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.

Table 5.3 Summary of Accommodations and Designated Supports Used by Students

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grade Level | Accessibility Resource | Resource Type | Students Assigned | Students Used |
| 5 | Embedded American Sign Language | ACC | 24 | 13 |
| 5 | Non-Embedded Print on Demand | ACC | 26 | 5 |
| 5 | Embedded Masking | DS | 1,047 | 34 |
| 5 | Embedded Text-to-Speech (English TTS) | DS | 4,969 | 2,660 |
| 8 | Embedded American Sign Language | ACC | 14 | 10 |
| 8 | Non-Embedded Print on Demand | ACC | 15 | 2 |
| 8 | Embedded Masking | DS | 1,527 | 62 |
| 8 | Embedded Text-to-Speech (English TTS) | DS | 4,789 | 1,549 |
| 10 | Embedded American Sign Language | ACC | 0 | 0 |
| 10 | Non-Embedded Print on Demand | ACC | 0 | 0 |
| 10 | Embedded Masking | DS | 9 | 0 |
| 10 | Embedded Text-to-Speech (English TTS) | DS | 74 | 18 |
| 11 | Embedded American Sign Language | ACC | 11 | 2 |
| 11 | Non-Embedded Print on Demand | ACC | 10 | 1 |
| 11 | Embedded Masking | DS | 152 | 2 |
| 11 | Embedded Text-to-Speech (English TTS) | DS | 916 | 185 |
| 12 | Embedded American Sign Language | ACC | 8 | 5 |
| 12 | Non-Embedded Print on Demand | ACC | 9 | 3 |
| 12 | Embedded Masking | DS | 2,255 | 24 |
| 12 | Embedded Text-to-Speech (English TTS) | DS | 3,760 | 400 |
| HS | Embedded American Sign Language | ACC | 19 | 7 |
| HS | Non-Embedded Print on Demand | ACC | 19 | 4 |
| HS | Embedded Masking | DS | 2,416 | 26 |
| HS | Embedded Text-to-Speech (English TTS) | DS | 4,750 | 603 |

### Practice and Training Tests

Practice and training tests are available publicly for the CAST. The practice tests simulated the experience of the computer-based CAST, and the training tests introduced students to the type of thinking needed to answer CAST items. Practice and training tests are available through the Practice and Training Test website linked on the Online Practice and Training Tests Portal web page on the CAASPP website.

The practice test was designed to be a more authentic representation of the summative CAST. It contained the same types of items and content as the summative CAST, as it was built using the blueprint. Practice items were periodically updated to meet accessibility standards and item type updates. Similar to the summative CAST, the practice test should take about two hours to be administered. *Practice Items Scoring Guides* for each grade level (CDE, 2020b; CDE, 2020c; CDE, 2020d) and *Practice Test Constructed Response Annotated Examples* were also available. 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.

The grade-level-specific training tests could be taken by students in all tested grades. All unique item types available on the operational test were covered in the training tests.

### Test Security and Confidentiality

For the 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 tests, including CDE staff, ETS staff, ETS subcontractors, LEA assessment coordinators, school assessment coordinators, students, parents/guardians, teachers, and cooperative educational service agency staff. ETS’ Code of Ethics required that all test information, including tangible materials (e.g., test items), confidential files (e.g., those containing personally identifiable student information), and processes related to test administration (e.g., the configuration of secure servers) were kept secure. To ensure security for all tests that ETS develops or handles, ETS maintains an Office of Testing Integrity (OTI).

All tests 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.

#### ETS’ Office of Testing Integrity

The OTI is a division of ETS that provides quality-assurance services for all ETS-managed testing programs. 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* (ETS, 2014), which supports the OTI’s 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 to help the public and auditors evaluate those products and services.

The OTI’s mission is to

* minimize any testing security violations that can impact the fairness of testing,
* minimize and investigate any security breach 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 balanced way that reflects the laws and professional standards governing the integrity of testing.

In an effort to enforce secure testing practices, the OTI strives to safeguard the various processes involved in a test development and administration cycle. For the CAST, those processes included the following:

* Test 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, PPTs, and student data—before, during, and after each test administration. For the CAST, as well as for all CAASPP assessments, the LEA CAASPP coordinator is responsible for keeping all electronic test materials secure, keeping student information confidential, and making sure the CAASPP test site coordinators and test administrators were properly trained regarding security policies and procedures.

The CAASPP test site coordinator is responsible for mitigating test security incidents at the test site and for reporting incidents to the LEA CAASPP coordinator. If the test site administered PPTs, the CAASPP test site coordinator is also responsible for the return of any secure materials to the LEA CAASPP coordinator, who, in turn, is responsible for returning any materials to the Scoring and Processing Center.

The test administrator is responsible for reporting testing incidents to the CAASPP test site coordinator and securely destroying printed and digital media for CAST items generated by the print-on-demand feature of the TDS (CDE, 2021b and 2021e).

The following measures ensured the security of CAASPP System assessments administered in 2020–2021:

* LEA CAASPP coordinators and test site coordinators must have electronically signed and submitted a “CAASPP Test Security Agreement for LEA CAASPP coordinators and CAASPP test site coordinators” form in TOMS before ETS granted the coordinators access to TOMS (5 *CCR,* Education, Division 1, Chapter 2, Subchapter 3.75, Article 1, Section 859[a]).
* 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]).

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 CAASPP test site 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 occur. Typically, ETS would perform site visits and testing procedure audits during the testing window; however, these visits were not made during the 2020–2021 CAASPP administration because many schools and LEAs were not open for in-person instruction as a result of the COVID-19 pandemic. It is expected that these visits will resume in future administrations per state health and safety guidelines. However, selected LEAs were audited remotely, with LEA CAASPP coordinators or other LEA staff responding to a series of questions about test administration.

#### 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 [*1.10 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 currently 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 2020–2021 CAASPP, 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 it had been posted.

#### Data Management in the Secure Database

ETS currently 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 operational 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, 2021d).

#### 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 (e.g., the values of all variables are as expected) before passing files to the ETS statistical analysis group (refer to section [*8.6 Quality Control of Psychometric Processes*](#_Quality_Control_of) for data validation processes undertaken by ETS Data Quality Services). The statistical analysis staff store the files on secure servers. All staff involved with the data adhere to the ETS Code of Ethics and the ETS Information Protection Policies to prevent any unauthorized access to the data.

#### Student Confidentiality

To meet 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 the CAST:

* Preliminary individual student reports for computer-based assessments in the California Educator Reporting System (CERS)
* Preliminary individual student reports for PPTs in CERS
* Individual SSRs (electronic)
* Internet reports—available on a public web reporting site—aggregated by state, county, LEA, or test site

##### Security of Results Files

ETS takes measures to protect files and reports that show students’ scores and achievement 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. ETS staff access to production databases is limited to personnel with a business need to access the data. User IDs for production systems must be person-specific or for systems use only.

ETS has implemented network controls for routers, gateways, switches, firewalls, network tier management, and network connectivity. Routers, gateways, and switches represent points of access between networks. However, these do not contain mass storage or represent points of vulnerability, particularly 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 the BS 25999-2 standard for business continuity and conducts disaster recovery exercises annually. ETS routinely backs up all data to either disks through deduplication or to tapes, all of which are stored off site.

Access to the ETS Computer Processing Center is controlled by employee and visitor identification badges. The Center is secured by doors that can be unlocked only by the badges of personnel who have functional responsibilities within its secure perimeter. Authorized personnel accompany visitors to the 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 events:

* Scoring
* Transfer of scores by means of secure data exchange
* Reporting
* Posting of aggregated data
* Storage

In addition to protecting the confidentiality of testing materials, 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 a test developed by ETS (e.g., a CAASPP assessment). 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 tests, which, in turn, compromise the reliability and validity of test results (CDE, 2021b). Whether intentional or unintentional, failure by staff or students to comply with security rules constitutes a test security incident. Test security incidents have impacts on scoring and affect students’ performance on the test.

LEA CAASPP coordinators and CAASPP test site coordinators ensured that all test security and summative 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 test. If an Appeal to a student’s test 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 CAASPP test site coordinator). The STAIRScase in TOMS provided the LEA CAASPP coordinator, the CDE, and CalTAC with the opportunity to interact and communicate regarding the STAIRS process (CDE, 2021b).

Prior to the assessment administration, ETS and the CDE agreed that the following types of STAIRS cases were also forwarded to the CDE:

* Student cheating or accessing unauthorized devices
* Security breach (where a student exposed secure materials)
* Student unable to review previous answers (i.e., 20-minute pause rule)

Appeals requests were reviewed by the CDE or CalTAC. When a request to submit an Appeal was approved, the coordinator received a system-generated email with the Appeal type that was approved (CDE, 2021b).

Types of Appeals available during the 2020–2021 CAASPP administration are described in table 5.4.

Table 5.4 Types of Appeals

|  |  |
| --- | --- |
| Type of Appeal | Description |
| Reset | Resetting a student’s summative assessment removed that assessment from the system and enabled the student to start a new assessment from the beginning. |
| Invalidate | Invalidated summative 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. The score was counted as “not proficient” for aggregation into the CAASPP results. |
| Re-open | Reopening a summative assessment allowed a student to access an assessment that had already been submitted or had expired. |
| Restore | Restoring a summative assessment returned an assessment from the Reset status to its prior status. This action could be performed only on tests that were reset previously. |
| Grace Period Extension | Permitting a grace period extension allowed the student to review previously answered questions 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 test, test security, or test validity. An impropriety can be corrected and contained at a local level. An impropriety should have been reported to the LEA CAASPP coordinator and CAASPP test site coordinator immediately. The coordinator reported 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 test, or impact test security or test validity. These circumstances can be corrected and contained at the local level and submitted using the STAIRS/Appeals process in TOMS. An irregularity should have been reported to the LEA CAASPP coordinator and CAASPP test site coordinator immediately. The coordinator must report the irregularity within 24 hours, using the STAIRS/Appeals process in TOMS.

##### Breach

A testing breach is an event that poses a threat to the validity of the test and requires 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 telephone. Following the call, the CAASPP test site coordinator or LEA CAASPP coordinator must report the incident using the STAIRS/Appeals process in TOMS within 24 hours. All other breaches were to be entered into STAIRS directly.

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 CDE decision to remove the test item(s) from the available secure item bank.

#### 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 CalTAC.

In most instances, an Appeal was submitted to address a test security breach or irregularity. The LEA CAASPP coordinator or CAASPP test site coordinator submitted Appeals in TOMS. All submitted Appeals were available for retrieval and review by the appropriate credentialed users within a given organization. However, the view of Appeals was restricted according to the user role as established in TOMS (CDE, 2021b). An Appeal could be requested only by the LEA CAASPP coordinator or CAASPP test site coordinator if prompted while filing a STAIRS case in TOMS (CDE, 2021e). Types of Appeals available during the 2020–2021 CAASPP administration are described in table 5.4.

Table 5.5 presents the number of Appeals approved and rejected in STAIRS in the 2020–‍2021 administration of CAST, for all grade levels combined.

Table 5.5 Number of Appeals Requested in STAIRS for the 2020–2021 Administration—All Grade Levels

|  |  |  |
| --- | --- | --- |
| Appeal Type | Number of Appeals Approved | Number of Appeals Rejected |
| Reset | 101 | 35 |
| Invalidate | 21 | 12 |
| Re-open | 358 | 0 |
| Restore | 0 | 0 |
| Grace Period Extension | 96 | 12 |

Table 5.6 presents the number and types of testing issues submitted in STAIRS. The most frequently reported incident was expired or accidentally submitted tests.

Table 5.6 Number and Types of Incidents Submitted in STAIRS for the 2020–2021 Administration—All Grade Levels

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Description | Appeal Type | Number of Incidents | Total Number of Statewide Student Identifiers (SSIDs) Submitted | Appeals SSIDs Approved |
| Accessibility Issue | Reset | 42 | 52 | 45 |
| Administered Incorrect Assessment | Reset, Re-open, or No Appeal | 35 | 115 | 69 |
| Administration Error | No Appeal | 8 | 0 | 0 |
| Data Entry Issue | Reset, Re-open, Invalidate, or No Appeal | 0 | 0 | 0 |
| Expired or Accidentally Submitted Test | Re-open | 67 | 344 | 343 |
| Exposing Secure Materials | Invalidate or No Appeal | 3 | 1 | 1 |
| Incorrect SSID Used | Reset or No Appeal | 3 | 3 | 2 |
| Restore from Reset | Restore | 0 | 0 | 0 |
| Student Cheating or Accessing Unauthorized Devices | Invalidate | 18 | 21 | 16 |
| Student Disruption | No Appeal | 0 | 0 | 0 |
| Technical Issues | Grace Period Extension or No Appeal | 24 | 108 | 96 |
| Validity Issue | Invalidate or Reset | 5 | 11 | 4 |

### 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 Department of Education. (2019). *Science test administration for high school students.* Sacramento, CA: California Department of Education.

California Department of Education. (2020a). *California assessment accessibility resources matrix*. Sacramento, CA: California Department of Education.

California Department of Education. (2020b). *California Science Test practice items scoring guide for grade eight*. Sacramento, CA: California Department of Education.

California Department of Education. (2020c). *California Science Test practice items scoring guide for grade five*. Sacramento, CA: California Department of Education.

California Department of Education. (2020d). *California Science Test practice items scoring guide for high school*. Sacramento, CA: California Department of Education.

California Department of Education. (2021a). *CAASPP and ELPAC accessibility guide for online testing*. Sacramento, CA: California Department of Education.

California Department of Education. (2021b). *CAASPP and ELPAC security incidents and Appeals procedure guide.* Sacramento, CA: California Department of Education.

California Department of Education. (2021c). *CAASPP and ELPAC technical specifications and configuration guide for online testing*. Sacramento, CA: California Department of Education.

California Department of Education. (2021d). *CAASPP and ELPAC Test Operations Management System user guide*. Sacramento, CA: California Department of Education.

California Department of Education. (2021e). *CAASPP online test administration manual.* Sacramento, CA: California Department of Education.

California Department of Education. (2021f). *Spring administration information for educators.* Sacramento, CA: California Department of Education.

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

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. (2020). *Smarter Balanced Assessment Consortium: Usability, accessibility, and accommodations guidelines.* Los Angeles: Smarter Balanced Assessment Consortium.

### Appendix 5.A: Test-Taking Rates

**Notes:**

* This set of tables shows the percent of test takers of selected demographic student groups in each test.
* The total numbers of registered students are derived from version 2 of the production data file (“P2”).
* 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 5.A. CAST Test-Taking Rates for Grade Five by Student Group

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Eligible Students | Number of Test Takers | Percent of Test Taking |
| All students | 442,327 | 62,554 | 14.1 |
| Male | 225,755 | 31,735 | 14.1 |
| Female | 216,539 | 30,817 | 14.2 |
| Nonbinary | 33 | 2 | 6.1 |
| EL | 84,498 | 11,428 | 13.5 |
| English only | 262,732 | 39,015 | 14.8 |
| RFEP | 77,650 | 9,690 | 12.5 |
| IFEP | 17,175 | 2,380 | 13.9 |
| ADEL | 0 | 0 | 0.0 |
| To be determined | 131 | 12 | 9.2 |
| English proficiency unknown | 141 | 29 | 20.6 |
| Economically disadvantaged | 265,350 | 34,356 | 12.9 |
| Not economically disadvantaged | 176,977 | 28,198 | 15.9 |
| American Indian or Alaska Native | 1,921 | 504 | 26.2 |
| Asian | 41,962 | 7,124 | 17.0 |
| Native Hawaiian or Other Pacific Islander | 1,896 | 237 | 12.5 |
| Filipino | 8,908 | 1,129 | 12.7 |
| Hispanic or Latino | 246,896 | 30,461 | 12.3 |
| Black or African American | 23,494 | 2,281 | 9.7 |
| White | 93,960 | 17,198 | 18.3 |
| Two or more races | 23,290 | 3,620 | 15.5 |
| Special education services | 56,656 | 7,086 | 12.5 |
| No special education services | 385,671 | 55,468 | 14.4 |
| Migrant education | 3,653 | 630 | 17.2 |
| Not migrant education | 438,674 | 61,924 | 14.1 |
| Military | 5,518 | 688 | 12.5 |
| Not military | 436,809 | 61,866 | 14.2 |
| Homeless | 15,974 | 1,479 | 9.3 |
| Not homeless | 426,353 | 61,075 | 14.3 |

Table 5.A. CAST Test-Taking Rates for Grade Eight by Student Group

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Eligible Students | Number of Test Takers | Percent of Test Taking |
| All students | 460,658 | 76,535 | 16.6 |
| Male | 236,337 | 39,078 | 16.5 |
| Female | 224,155 | 37,433 | 16.7 |
| Nonbinary | 166 | 24 | 14.5 |
| EL | 63,386 | 10,933 | 17.2 |
| English only | 259,661 | 43,227 | 16.6 |
| RFEP | 118,454 | 18,934 | 16.0 |
| IFEP | 18,911 | 3,404 | 18.0 |
| ADEL | 0 | 0 | 0.0 |
| To be determined | 128 | 6 | 4.7 |
| English proficiency unknown | 118 | 31 | 26.3 |
| Economically disadvantaged | 272,257 | 42,812 | 15.7 |
| Not economically disadvantaged | 188,401 | 33,723 | 17.9 |
| American Indian or Alaska Native | 2,179 | 560 | 25.7 |
| Asian | 42,971 | 7,880 | 18.3 |
| Native Hawaiian or Other Pacific Islander | 2,130 | 277 | 13.0 |
| Filipino | 9,630 | 1,830 | 19.0 |
| Hispanic or Latino | 257,240 | 39,538 | 15.4 |
| Black or African American | 24,341 | 2,777 | 11.4 |
| White | 100,253 | 19,830 | 19.8 |
| Two or more races | 21,914 | 3,843 | 17.5 |
| Special education services | 56,858 | 8,390 | 14.8 |
| No special education services | 403,800 | 68,145 | 16.9 |
| Migrant education | 3,480 | 681 | 19.6 |
| Not migrant education | 457,178 | 75,854 | 16.6 |
| Military | 5,586 | 1,068 | 19.1 |
| Not military | 455,072 | 75,467 | 16.6 |
| Homeless | 14,696 | 1,825 | 12.4 |
| Not homeless | 445,962 | 74,710 | 16.8 |

Table 5.A. CAST Test-Taking Rates for Grade Ten by Student Group

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Eligible Students | Number of Test Takers | Percent of Test Taking |
| All students | 7,609 | 2,281 | 30.0 |
| Male | 3,802 | 1,143 | 30.1 |
| Female | 3,799 | 1,137 | 29.9 |
| Nonbinary | 8 | 1 | 12.5 |
| EL | 449 | 158 | 35.2 |
| English only | 5,513 | 1,454 | 26.4 |
| RFEP | 1,312 | 522 | 39.8 |
| IFEP | 332 | 147 | 44.3 |
| ADEL | 0 | 0 | 0.0 |
| To be determined | 0 | 0 | 0.0 |
| English proficiency unknown | 3 | 0 | 0.0 |
| Economically disadvantaged | 2,462 | 1,000 | 40.6 |
| Not economically disadvantaged | 5,147 | 1,281 | 24.9 |
| American Indian or Alaska Native | 68 | 22 | 32.4 |
| Asian | 873 | 324 | 37.1 |
| Native Hawaiian or Other Pacific Islander | 37 | 7 | 18.9 |
| Filipino | 154 | 19 | 12.3 |
| Hispanic or Latino | 2,402 | 895 | 37.3 |
| Black or African American | 281 | 53 | 18.9 |
| White | 3,310 | 832 | 25.1 |
| Two or more races | 484 | 129 | 26.7 |
| Special education services | 1,172 | 229 | 19.5 |
| No special education services | 6,437 | 2,052 | 31.9 |
| Migrant education | 26 | 14 | 53.8 |
| Not migrant education | 7,583 | 2,267 | 29.9 |
| Military | 35 | 14 | 40.0 |
| Not military | 7,574 | 2,267 | 29.9 |
| Homeless | 72 | 28 | 38.9 |
| Not homeless | 7,537 | 2,253 | 29.9 |

Table 5.A. CAST Test-Taking Rates for Grade Eleven by Student Group

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Eligible Students | Number of Test Takers | Percent of Test Taking |
| All students | 177,234 | 44,207 | 24.9 |
| Male | 89,668 | 22,070 | 24.6 |
| Female | 87,516 | 22,118 | 25.3 |
| Nonbinary | 50 | 19 | 38.0 |
| EL | 16,513 | 3,477 | 21.1 |
| English only | 93,458 | 24,341 | 26.0 |
| RFEP | 57,789 | 14,496 | 25.1 |
| IFEP | 9,435 | 1,884 | 20.0 |
| ADEL | 2 | 0 | 0.0 |
| To be determined | 27 | 1 | 3.7 |
| English proficiency unknown | 10 | 8 | 80.0 |
| Economically disadvantaged | 105,599 | 24,109 | 22.8 |
| Not economically disadvantaged | 71,635 | 20,098 | 28.1 |
| American Indian or Alaska Native | 861 | 304 | 35.3 |
| Asian | 16,282 | 4,511 | 27.7 |
| Native Hawaiian or Other Pacific Islander | 755 | 178 | 23.6 |
| Filipino | 4,917 | 1,327 | 27.0 |
| Hispanic or Latino | 98,678 | 21,898 | 22.2 |
| Black or African American | 8,817 | 1,431 | 16.2 |
| White | 40,353 | 12,768 | 31.6 |
| Two or more races | 6,571 | 1,790 | 27.2 |
| Special education services | 19,192 | 3,807 | 19.8 |
| No special education services | 158,042 | 40,400 | 25.6 |
| Migrant education | 813 | 271 | 33.3 |
| Not migrant education | 176,421 | 43,936 | 24.9 |
| Military | 2,600 | 658 | 25.3 |
| Not military | 174,634 | 43,549 | 24.9 |
| Homeless | 4,364 | 1,394 | 31.9 |
| Not homeless | 172,870 | 42,813 | 24.8 |

Table 5.A. CAST Test-Taking Rates for Grade Twelve by Student Group

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Eligible Students | Number of Test Takers | Percent of Test Taking |
| All students | 410,324 | 62,218 | 15.2 |
| Male | 208,290 | 30,662 | 14.7 |
| Female | 201,902 | 31,534 | 15.6 |
| Nonbinary | 132 | 22 | 16.7 |
| EL | 34,312 | 5,171 | 15.1 |
| English only | 217,485 | 31,175 | 14.3 |
| RFEP | 135,865 | 22,692 | 16.7 |
| IFEP | 22,174 | 3,101 | 14.0 |
| ADEL | 313 | 69 | 22.0 |
| To be determined | 81 | 3 | 3.7 |
| English proficiency unknown | 94 | 7 | 7.4 |
| Economically disadvantaged | 219,426 | 37,401 | 17.0 |
| Not economically disadvantaged | 190,898 | 24,817 | 13.0 |
| American Indian or Alaska Native | 1,845 | 324 | 17.6 |
| Asian | 41,486 | 5,329 | 12.8 |
| Native Hawaiian or Other Pacific Islander | 2,034 | 244 | 12.0 |
| Filipino | 11,740 | 1,929 | 16.4 |
| Hispanic or Latino | 221,718 | 36,197 | 16.3 |
| Black or African American | 21,031 | 2,449 | 11.6 |
| White | 93,996 | 13,483 | 14.3 |
| Two or more races | 16,474 | 2,263 | 13.7 |
| Special education services | 44,142 | 5,739 | 13.0 |
| No special education services | 366,182 | 56,479 | 15.4 |
| Migrant education | 2,694 | 479 | 17.8 |
| Not migrant education | 407,630 | 61,739 | 15.1 |
| Military | 5,922 | 967 | 16.3 |
| Not military | 404,402 | 61,251 | 15.1 |
| Homeless | 12,514 | 2,163 | 17.3 |
| Not homeless | 397,810 | 60,055 | 15.1 |

Table 5.A. CAST Test-Taking Rates for High School (All Grades Tested) by Student Group

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Eligible Students | Number of Test Takers | Percent of Test Taking |
| All students | 595,167 | 108,706 | 18.3 |
| Male | 301,760 | 53,875 | 17.9 |
| Female | 293,217 | 54,789 | 18.7 |
| Nonbinary | 190 | 42 | 22.1 |
| EL | 51,274 | 8,806 | 17.2 |
| English only | 316,456 | 56,970 | 18.0 |
| RFEP | 194,966 | 37,710 | 19.3 |
| IFEP | 31,941 | 5,132 | 16.1 |
| ADEL | 315 | 69 | 21.9 |
| To be determined | 108 | 4 | 3.7 |
| English proficiency unknown | 107 | 15 | 14.0 |
| Economically disadvantaged | 327,487 | 62,510 | 19.1 |
| Not economically disadvantaged | 267,680 | 46,196 | 17.3 |
| American Indian or Alaska Native | 2,774 | 650 | 23.4 |
| Asian | 58,641 | 10,164 | 17.3 |
| Native Hawaiian or Other Pacific Islander | 2,826 | 429 | 15.2 |
| Filipino | 16,811 | 3,275 | 19.5 |
| Hispanic or Latino | 322,798 | 58,990 | 18.3 |
| Black or African American | 30,129 | 3,933 | 13.1 |
| White | 137,659 | 27,083 | 19.7 |
| Two or more races | 23,529 | 4,182 | 17.8 |
| Special education services | 64,506 | 9,775 | 15.2 |
| No special education services | 530,661 | 98,931 | 18.6 |
| Migrant education | 3,533 | 764 | 21.6 |
| Not migrant education | 591,634 | 107,942 | 18.2 |
| Military | 8,557 | 1,639 | 19.2 |
| Not military | 586,610 | 107,067 | 18.3 |
| Homeless | 16,950 | 3,585 | 21.2 |
| Not homeless | 578,217 | 105,121 | 18.2 |

### Appendix 5.B: Demographic Summary

**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 the CAST in grades five and eight. High school students are required to test once in grade ten, eleven, or twelve. Because the 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 5.B.1 through table 5.B.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 due to rounding. In addition, the percentages of students within a race or ethnicity category may not sum to 100 due to missing race or ethnicity data for some students.

Table 5.B. Demographic Summary for Grade Five

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Valid Scores | Percent of Valid Scores | Population Percent |
| All students | 62,238 | 100.0 | 100.0 |
| Male | 31,577 | 50.7 | 51.2 |
| Female | 30,659 | 49.3 | 48.8 |
| Nonbinary | 2 | 0.0 | 0.0 |
| EL | 11,335 | 18.2 | 19.3 |
| English only | 38,862 | 62.4 | 59.4 |
| RFEP | 9,628 | 15.5 | 17.3 |
| IFEP | 2,372 | 3.8 | 3.8 |
| ADEL | 0 | 0.0 | 0.0 |
| To be determined | 12 | 0.0 | 0.0 |
| English proficiency unknown | 29 | 0.0 | 0.0 |
| Economically disadvantaged | 34,089 | 54.8 | 61.2 |
| Not economically disadvantaged | 28,149 | 45.2 | 38.8 |

Table 5.B.1 *(continuation one)*

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Valid Scores | Percent of Valid Scores | Population Percent |
| American Indian or Alaska Native (All) | 503 | 0.8 | 0.5 |
| Asian (All) | 7,115 | 11.4 | 9.4 |
| Native Hawaiian or Other Pacific Islander (All) | 237 | 0.4 | 0.4 |
| Filipino (All) | 1,128 | 1.8 | 2.0 |
| Hispanic or Latino (All) | 30,227 | 48.6 | 55.8 |
| Black or African American (All) | 2,256 | 3.6 | 5.3 |
| White (All) | 17,170 | 27.6 | 21.3 |
| Two or more races (All) | 3,602 | 5.8 | 4.5 |
| Special education services | 7,041 | 11.3 | 13.7 |
| No special education services | 55,197 | 88.7 | 86.3 |
| Migrant education | 626 | 1.0 | 0.8 |
| Not migrant education | 61,612 | 99.0 | 99.2 |
| Military | 688 | 1.1 | 1.2 |
| Not military | 61,550 | 98.9 | 98.8 |
| Homeless | 1,463 | 2.4 | 3.6 |
| Not homeless | 60,775 | 97.6 | 96.4 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 147 | 0.2 | 0.1 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 356 | 0.6 | 0.3 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 5,451 | 8.8 | 6.1 |
| Asian (Primary ethnicity—Economically disadvantaged) | 1,664 | 2.7 | 3.4 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 100 | 0.2 | 0.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 137 | 0.2 | 0.3 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 785 | 1.3 | 1.3 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 343 | 0.6 | 0.7 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 7,330 | 11.8 | 12.1 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 22,897 | 36.8 | 43.6 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 682 | 1.1 | 1.3 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 1,574 | 2.5 | 4.1 |

Table 5.B.1 *(continuation two)*

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Valid Scores | Percent of Valid Scores | Population Percent |
| White (Primary ethnicity—Not economically disadvantaged) | 11,348 | 18.2 | 14.4 |
| White (Primary ethnicity—Economically disadvantaged) | 5,822 | 9.4 | 6.9 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 2,306 | 3.7 | 2.9 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 1,296 | 2.1 | 1.6 |

Table 5.B. Demographic Summary for Grade Eight

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Valid Scores | Percent of Valid Scores | Population Percent |
| All students | 75,000 | 100.0 | 100.0 |
| Male | 38,344 | 51.1 | 51.5 |
| Female | 36,633 | 48.8 | 48.5 |
| Nonbinary | 23 | 0.0 | 0.0 |
| EL | 10,521 | 14.0 | 14.0 |
| English only | 42,628 | 56.8 | 56.3 |
| RFEP | 18,470 | 24.6 | 25.5 |
| IFEP | 3,344 | 4.5 | 4.1 |
| ADEL | 0 | 0.0 | 0.0 |
| To be determined | 6 | 0.0 | 0.0 |
| English proficiency unknown | 31 | 0.0 | 0.0 |
| Economically disadvantaged | 41,609 | 55.5 | 60.3 |
| Not economically disadvantaged | 33,391 | 44.5 | 39.7 |
| American Indian or Alaska Native (All) | 553 | 0.7 | 0.5 |
| Asian (All) | 7,826 | 10.4 | 9.3 |
| Native Hawaiian or Other Pacific Islander (All) | 268 | 0.4 | 0.5 |
| Filipino (All) | 1,807 | 2.4 | 2.1 |
| Hispanic or Latino (All) | 38,340 | 51.1 | 55.9 |
| Black or African American (All) | 2,717 | 3.6 | 5.3 |
| White (All) | 19,707 | 26.3 | 21.7 |
| Two or more races (All) | 3,782 | 5.0 | 4.1 |
| Special education services | 8,188 | 10.9 | 13.2 |
| No special education services | 66,812 | 89.1 | 86.8 |
| Migrant education | 668 | 0.9 | 0.8 |
| Not migrant education | 74,332 | 99.1 | 99.2 |
| Military | 1,049 | 1.4 | 1.2 |
| Not military | 73,951 | 98.6 | 98.8 |
| Homeless | 1,773 | 2.4 | 3.1 |
| Not homeless | 73,227 | 97.6 | 96.9 |

Table 5.B.2 *(continuation)*

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Valid Scores | Percent of Valid Scores | Population Percent |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 182 | 0.2 | 0.2 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 371 | 0.5 | 0.3 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 5,816 | 7.8 | 5.9 |
| Asian (Primary ethnicity—Economically disadvantaged) | 2,010 | 2.7 | 3.4 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 119 | 0.2 | 0.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 149 | 0.2 | 0.3 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 1,233 | 1.6 | 1.4 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 574 | 0.8 | 0.7 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 9,276 | 12.4 | 12.5 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 29,064 | 38.8 | 43.4 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 868 | 1.2 | 1.4 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 1,849 | 2.5 | 3.8 |
| White (Primary ethnicity—Not economically disadvantaged) | 13,428 | 17.9 | 15.1 |
| White (Primary ethnicity—Economically disadvantaged) | 6,279 | 8.4 | 6.7 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 2,469 | 3.3 | 2.7 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 1,313 | 1.8 | 1.4 |

Table 5.B. Demographic Summary for Grade Ten

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Valid Scores | Percent of Valid Scores | Population Percent |
| All students | 2,270 | 100.0 | 100.0 |
| Male | 1,135 | 50.0 | 49.5 |
| Female | 1,134 | 50.0 | 50.1 |
| Nonbinary | 1 | 0.0 | 0.1 |
| EL | 156 | 6.9 | 5.8 |
| English only | 1,448 | 63.8 | 71.9 |
| RFEP | 519 | 22.9 | 17.6 |
| IFEP | 147 | 6.5 | 4.4 |
| ADEL | 0 | 0.0 | 0.0 |
| To be determined | 0 | 0.0 | 0.0 |
| English proficiency unknown | 0 | 0.0 | 0.3 |
| Economically disadvantaged | 992 | 43.7 | 37.9 |
| Not economically disadvantaged | 1,278 | 56.3 | 61.8 |
| American Indian or Alaska Native (All) | 22 | 1.0 | 0.7 |
| Asian (All) | 323 | 14.2 | 11.5 |
| Native Hawaiian or Other Pacific Islander (All) | 7 | 0.3 | 0.5 |
| Filipino (All) | 19 | 0.8 | 2.0 |
| Hispanic or Latino (All) | 890 | 39.2 | 31.4 |
| Black or African American (All) | 53 | 2.3 | 3.8 |
| White (All) | 831 | 36.6 | 43.3 |
| Two or more races (All) | 125 | 5.5 | 5.1 |
| Special education services | 226 | 10.0 | 15.2 |
| No special education services | 2,044 | 90.0 | 84.6 |
| Migrant education | 13 | 0.6 | 0.3 |
| Not migrant education | 2,257 | 99.4 | 99.4 |
| Military | 14 | 0.6 | 0.5 |
| Not military | 2,256 | 99.4 | 99.3 |
| Homeless | 28 | 1.2 | 1.0 |
| Not homeless | 2,242 | 98.8 | 98.7 |

Table 5.B.3 *(continuation)*

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Valid Scores | Percent of Valid Scores | Population Percent |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 9 | 0.4 | 0.4 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 13 | 0.6 | 0.3 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 292 | 12.9 | 9.3 |
| Asian (Primary ethnicity—Economically disadvantaged) | 31 | 1.4 | 2.3 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 6 | 0.3 | 0.3 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 1 | 0.0 | 0.2 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 16 | 0.7 | 1.8 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 3 | 0.1 | 0.2 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 276 | 12.2 | 12.0 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 614 | 27.0 | 19.4 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 26 | 1.1 | 1.6 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 27 | 1.2 | 2.2 |
| White (Primary ethnicity—Not economically disadvantaged) | 567 | 25.0 | 32.1 |
| White (Primary ethnicity—Economically disadvantaged) | 264 | 11.6 | 11.2 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 86 | 3.8 | 3.5 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 39 | 1.7 | 1.6 |

Table 5.B. Demographic Summary for Grade Eleven

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Valid Scores | Percent of Valid Scores | Population Percent |
| All students | 43,791 | 100.0 | 100.0 |
| Male | 21,859 | 49.9 | 50.5 |
| Female | 21,913 | 50.0 | 49.3 |
| Nonbinary | 19 | 0.0 | 0.0 |
| EL | 3,402 | 7.8 | 9.5 |
| English only | 24,168 | 55.2 | 52.3 |
| RFEP | 14,341 | 32.7 | 32.7 |
| IFEP | 1,871 | 4.3 | 5.3 |
| ADEL | 0 | 0.0 | 0.0 |
| To be determined | 1 | 0.0 | 0.0 |
| English proficiency unknown | 8 | 0.0 | 0.2 |
| Economically disadvantaged | 23,793 | 54.3 | 60.1 |
| Not economically disadvantaged | 19,998 | 45.7 | 39.6 |
| American Indian or Alaska Native (All) | 304 | 0.7 | 0.5 |
| Asian (All) | 4,494 | 10.3 | 9.2 |
| Native Hawaiian or Other Pacific Islander (All) | 174 | 0.4 | 0.4 |
| Filipino (All) | 1,319 | 3.0 | 2.8 |
| Hispanic or Latino (All) | 21,606 | 49.3 | 55.8 |
| Black or African American (All) | 1,402 | 3.2 | 5.0 |
| White (All) | 12,710 | 29.0 | 22.4 |
| Two or more races (All) | 1,782 | 4.1 | 3.2 |
| Special education services | 3,724 | 8.5 | 10.6 |
| No special education services | 40,067 | 91.5 | 89.2 |
| Migrant education | 266 | 0.6 | 0.5 |
| Not migrant education | 43,525 | 99.4 | 99.3 |
| Military | 654 | 1.5 | 1.5 |
| Not military | 43,137 | 98.5 | 98.3 |
| Homeless | 1,375 | 3.1 | 2.5 |
| Not homeless | 42,416 | 96.9 | 97.3 |

Table 5.B.4 *(continuation)*

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Valid Scores | Percent of Valid Scores | Population Percent |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 139 | 0.3 | 0.2 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 165 | 0.4 | 0.3 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 2,749 | 6.3 | 4.9 |
| Asian (Primary ethnicity—Economically disadvantaged) | 1,745 | 4.0 | 4.3 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 81 | 0.2 | 0.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 93 | 0.2 | 0.3 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 865 | 2.0 | 1.7 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 454 | 1.0 | 1.1 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 5,515 | 12.6 | 12.9 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 16,091 | 36.7 | 43.0 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 520 | 1.2 | 1.6 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 882 | 2.0 | 3.4 |
| White (Primary ethnicity—Not economically disadvantaged) | 8,982 | 20.5 | 15.8 |
| White (Primary ethnicity—Economically disadvantaged) | 3,728 | 8.5 | 6.6 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 1,147 | 2.6 | 2.1 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 635 | 1.5 | 1.1 |

Table 5.B. Demographic Summary for Grade Twelve

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Valid Scores | Percent of Valid Scores | Population Percent |
| All students | 61,723 | 100.0 | 100.0 |
| Male | 30,416 | 49.3 | 50.0 |
| Female | 31,285 | 50.7 | 48.7 |
| Nonbinary | 22 | 0.0 | 0.0 |
| EL | 5,086 | 8.2 | 8.2 |
| English only | 30,988 | 50.2 | 52.3 |
| RFEP | 22,485 | 36.4 | 32.7 |
| IFEP | 3,086 | 5.0 | 5.4 |
| ADEL | 69 | 0.1 | 0.1 |
| To be determined | 2 | 0.0 | 0.0 |
| English proficiency unknown | 7 | 0.0 | 1.3 |
| Economically disadvantaged | 37,023 | 60.0 | 55.7 |
| Not economically disadvantaged | 24,700 | 40.0 | 43.1 |
| American Indian or Alaska Native (All) | 321 | 0.5 | 0.5 |
| Asian (All) | 5,296 | 8.6 | 10.1 |
| Native Hawaiian or Other Pacific Islander (All) | 241 | 0.4 | 0.5 |
| Filipino (All) | 1,916 | 3.1 | 2.9 |
| Hispanic or Latino (All) | 35,846 | 58.1 | 53.2 |
| Black or African American (All) | 2,419 | 3.9 | 5.0 |
| White (All) | 13,436 | 21.8 | 22.6 |
| Two or more races (All) | 2,248 | 3.6 | 3.4 |
| Special education services | 5,660 | 9.2 | 10.6 |
| No special education services | 56,063 | 90.8 | 88.2 |
| Migrant education | 474 | 0.8 | 0.7 |
| Not migrant education | 61,249 | 99.2 | 98.1 |
| Military | 960 | 1.6 | 1.4 |
| Not military | 60,763 | 98.4 | 97.3 |
| Homeless | 2,148 | 3.5 | 3.1 |
| Not homeless | 59,575 | 96.5 | 95.7 |

Table 5.B.5 *(continuation)*

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Valid Scores | Percent of Valid Scores | Population Percent |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 116 | 0.2 | 0.2 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 205 | 0.3 | 0.3 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 3,199 | 5.2 | 5.9 |
| Asian (Primary ethnicity—Economically disadvantaged) | 2,097 | 3.4 | 4.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 109 | 0.2 | 0.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 132 | 0.2 | 0.3 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 1,197 | 1.9 | 1.9 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 719 | 1.2 | 1.0 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 8,204 | 13.3 | 13.8 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 27,642 | 44.8 | 39.5 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 815 | 1.3 | 1.7 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 1,604 | 2.6 | 3.3 |
| White (Primary ethnicity—Not economically disadvantaged) | 9,612 | 15.6 | 16.7 |
| White (Primary ethnicity—Economically disadvantaged) | 3,824 | 6.2 | 5.9 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 1,448 | 2.3 | 2.3 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 800 | 1.3 | 1.1 |

Table 5.B. Demographic Summary for High School (All Grades Tested)

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Valid Scores | Percent of Valid Scores | Population Percent |
| All students | 107,784 | 100.0 | 100.0 |
| Male | 53,410 | 49.6 | 50.2 |
| Female | 54,332 | 50.4 | 48.9 |
| Nonbinary | 42 | 0.0 | 0.0 |
| EL | 8,644 | 8.0 | 8.5 |
| English only | 56,604 | 52.5 | 52.6 |
| RFEP | 37,345 | 34.6 | 32.5 |
| IFEP | 5,104 | 4.7 | 5.4 |
| ADEL | 69 | 0.1 | 0.1 |
| To be determined | 3 | 0.0 | 0.0 |
| English proficiency unknown | 15 | 0.0 | 0.9 |
| Economically disadvantaged | 61,808 | 57.3 | 56.8 |
| Not economically disadvantaged | 45,976 | 42.7 | 42.3 |
| American Indian or Alaska Native (All) | 647 | 0.6 | 0.5 |
| Asian (All) | 10,113 | 9.4 | 9.8 |
| Native Hawaiian or Other Pacific Islander (All) | 422 | 0.4 | 0.5 |
| Filipino (All) | 3,254 | 3.0 | 2.8 |
| Hispanic or Latino (All) | 58,342 | 54.1 | 53.7 |
| Black or African American (All) | 3,874 | 3.6 | 5.0 |
| White (All) | 26,977 | 25.0 | 22.8 |
| Two or more races (All) | 4,155 | 3.9 | 3.4 |
| Special education services | 9,610 | 8.9 | 10.7 |
| No special education services | 98,174 | 91.1 | 88.4 |
| Migrant education | 753 | 0.7 | 0.6 |
| Not migrant education | 107,031 | 99.3 | 98.5 |
| Military | 1,628 | 1.5 | 1.4 |
| Not military | 106,156 | 98.5 | 97.6 |
| Homeless | 3,551 | 3.3 | 2.9 |
| Not homeless | 104,233 | 96.7 | 96.2 |

Table 5.B.6 *(continuation)*

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Number of Valid Scores | Percent of Valid Scores | Population Percent |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 264 | 0.2 | 0.2 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 383 | 0.4 | 0.3 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 6,240 | 5.8 | 5.7 |
| Asian (Primary ethnicity—Economically disadvantaged) | 3,873 | 3.6 | 4.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 196 | 0.2 | 0.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 226 | 0.2 | 0.3 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 2,078 | 1.9 | 1.8 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 1,176 | 1.1 | 1.0 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 13,995 | 13.0 | 13.5 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 44,347 | 41.1 | 40.2 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 1,361 | 1.3 | 1.7 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 2,513 | 2.3 | 3.3 |
| White (Primary ethnicity—Not economically disadvantaged) | 19,161 | 17.8 | 16.7 |
| White (Primary ethnicity—Economically disadvantaged) | 7,816 | 7.3 | 6.2 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 2,681 | 2.5 | 2.3 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 1,474 | 1.4 | 1.1 |

## Scoring and Reporting

To determine individual students’ scores for the California Science Test (CAST), student item responses were scored and individual student scores—overall scale scores and science domain level scores—were calculated based on 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 for the CAST, as well as the various types of scores and reports that were generated. This chapter also presents information on the concept of measurement error and how measurement error should be considered when interpreting student test scores.

### CAST Scoring Process

#### Scoring for Constructed-Response Items

While constructed-response (CR) items used in the operational segments (i.e., segments A and B) of the 2020–‍2021 CAST were scored, scoring for CR field test items was not conducted for grades five and eight because of low student volumes that did not support the ability to conduct range finding and psychometric analyses. The California State Board of Education received a waiver for the accountability of the statewide standardized assessments. Because of the waiver, the number of students taking the California Assessment of Student Performance and Progress (CAASPP) assessments was much smaller than the number of students in a typical year. Range finding and psychometric analyses were completed for data from the high school grade band, where student numbers were sufficient. Subsection [*6.1.1 Scoring for Constructed-Response Items*](#_Scoring_for_Constructed-Response) describes the scoring process implemented for standard test administrations.

The CAST at each tested grade level included selected-response (SR) and 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.

Out of 29 CR items used in segments A and B of the 2020–2021 operational forms across three grade levels, two of them were human-scored and 27 were AI-scored. The AI models were built and approved for operational use by means of either the data from the 2017–2018 field test administration or the 2018–2019 first operational 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 high school 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

During a typical administration, there are two CR item sampling processes for the 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 to ensure representation of demographic student groups are used.

###### Sampling Process for Double Scoring of Operational Constructed-Response Items

For the 2020–2021 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). ETS’ Psychometric Analysis & Research (PAR) group conducted an evaluation to ensure that the sample represented population demographics, including gender, ethnicity, special education services 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 blocks: two operational discrete blocks in Segment A, two operational performance tasks (PTs) in Segment B, and one discrete block or one PT in Segment C (field test segment).

Typically, for CR field test prompts in grades five and eight assessments, 1,800 responses per prompt are sampled. For high school, because students may test in grade ten, eleven, or twelve, the percentage of students from each grade taking the test could be drastically different. To account for the uncertainty in sampling composition by grade level, 2,500 responses per prompt are 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 not conducted for the 2020–2021 administration for grade five and grade eight because of the insufficient sample size collected after the waiver for accountability was approved in April 2021.

##### Human Scoring

###### Scoring Rubric Development

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 2020–2021 administration, range finding was conducted for the high school assessment only.

In a typical test administration, soon after receiving a sufficient volume of CR responses from California schools—approximately 500 responses per CR prompt to be field-tested—ETS begins the range finding process by randomly selecting a wide variety of student response samples for each field test prompt. The goal is to ensure sufficient responses at each score point on the rubric to create sets of responses for training and certifying (qualifying) raters (scorers) and for monitoring raters during the scoring process. Another part of the range finding process includes annotating responses to provide further guidance on why a response received a certain rating. The following steps describe how the range finding process is implemented:

1. ETS Assessment and Learning Technology Research & Development (ALTRD) staff use 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), apply a subsequent score to initially selected responses. Scored samples needed for various purposes are summarized in table 6.1.

Table 6.1 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 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 (60 unique prompts total) | 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 ten samples for each score point |

1. California educators are invited by SCOE to apply to participate in the 2020–2021 Range Finding Meeting. From the applicants, SCOE compiles 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 is sent to the CDE for approval. Thirty California educators are approved to participate, 10 per grade level or grade band, with a minimum of six per grade level or grade band with science teaching experience. (However, in the 2020–2021 administration, only 10 high school educators participated in range finding, because of low volumes for grade five and grade eight.)
2. ETS ALTRD staff train the educators on how to score the sample responses using the ONE scoring system. The 10 educators for each grade level or grade band are assigned to one of three groups. Each group is assigned a set of field test prompts for that grade level or grade band, and the educators in each group score the samples for each of their assigned prompts. The educators also complete a feedback survey after scoring the samples for each prompt. As a result, each sample response has a minimum of three educator scores.
3. The ALTRD science lead for each grade level or grade band reviews the educator scores and feedback. ALTRD staff compiles the written and verbal feedback from the educators and provides it to the CDE prior to follow-up virtual meetings with the educators. The purpose of the follow-up meetings is to discuss samples that receive discrepant scores from the educators. The ALTRD science leads identify samples with discrepant scores and facilitate a meeting with each group of educators to reconcile the discrepant scores. When scheduled, CDE staff observe these virtual meetings.
4. Based on the educator scores and feedback, ALTRD staff select responses to include in the required sets for each field test prompt (i.e., certification, benchmark, training, calibration, and validity, as summarized in table 6.1). Samples in the benchmark and training sets are 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 review the proposed benchmark sets for each prompt in the ONE system and approve the sets in reconciliation meetings with ETS. ALTRD staff then create all final sets in the ONE system and use 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 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 6.2 shows the characteristics of the CAST raters. Among the 325 raters scoring for the 2020–2021 CAST, 161 had teaching experience in science, 48 currently worked in a kindergarten (K)–‍12 school in California, 12 were fluent in Spanish, and 47 had experience teaching in a K–12 school.

Table 6.2 Summary of Characteristics of Human Raters Scoring the CAST

|  |  |
| --- | --- |
| Characteristic | N |
| Experience teaching in Science | 161 |
| Fluent in Spanish | 12 |
| Experience teaching in a K–12 school | 47 |
| Currently works in a K–12 school in California | 48 |
| Others—Not meeting any of the previous criteria | 57 |
| **Total raters scoring in 2020–2021** | **325** |

Certification served as an initial screening to ensure that 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 (i.e., a response that can earn 2, 1, or 0 points) selected from among the operational high school 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 high school prompt, the rater was eligible to calibrate on the grade-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 grade-level test prompt. If raters did not pass either certification set, they were excused from scoring the 2020–2021 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 provide rater feedback. ETS assessment specialists also conducted virtual training sessions for scoring leaders, using online conferencing tools to provide specific grade-level guidance on specific prompts.

###### Training for Raters

Training for raters occurred within the ONE system. Raters were provided 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 based on 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 2020–2021 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.

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 (e.g., January through July) to check scoring accuracy on a prescored set of responses. For the 2020–2021 CAST, raters were calibrated once for each CR item during a three-day scoring period.

For the 2020–2021 CAST administration, raters were permitted to score any prompt for a grade 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 10 percent of responses per item were double-scored (with a maximum of 1,800 responses) 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 40 percent of responses 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 [*6.1.1.1.2 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 be 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 ALTRD, 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 (i.e., interrater reliability).

###### Interrater Reliability for Operational Items

The ONE system captured interrater reliability by monitoring data for responses that were double-scored. Approximately 9,600 CAST responses per item were double-scored. The statistics included the percentage agreement between the two raters, kappa, and the quadratic-weighted kappa (QWK). For detailed descriptions of these statistics, refer to subsection [*7.6.2.1 Interrater Agreement*](#_Interrater_Agreement_1) in [*Chapter 7: Psychometric Analyses*](#_Chapter_7:_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.

The interrater reliability statistics are shown in table 6.3 through table 6.6. Table 6.3 through table 6.5 include the operational items that were scored by AI and by human raters, where Rater 1 scores in these tables refer to the AI scores. Table 6.6 includes the operational items that were scored by two human raters.

These tables show that the percentage of students for whom the human raters and the AI raters were in exact agreement ranged from 78.86 percent to 95.44 percent for 1-point items and 69.78 percent to 96.09 percent for 2-point items. The percentage of students for whom the human raters were in exact agreement was 86.61 percent for the 1-point items and 79.40 for the 2-point items.

These tables also showed that the QWK ranged from 0.58 to 0.96 for AI-scored items and 0.62 to 0.78 for human-scored items, which indicates a moderate-to-high level of agreement between two raters.

Table 6.3 Interrater Reliability and Descriptive Statistics for the Ratings by AI and Human Raters in AI-Scoring of Operational Items for Grade Five

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Prompt | Item ID | Score Points | Rater 1 N | Rater 2 N | Kappa | QWK | Percent Exact | Percent Adjacent | Percent Exact + Adjacent | Rater 1 Item Score Mean | Rater 1 Item Score Standard Deviation (SD) | Rater 2 Item Score Mean | Rater 2 Item Score SD |
| 1 | VH668026 | 1 | 984 | 984 | 0.58 | 0.58 | 78.86 | 21.14 | 100.00 | 0.36 | 0.48 | 0.52 | 0.50 |
| 2 | VH831833 | 2 | 1,691 | 1,691 | 0.79 | 0.88 | 86.04 | 13.66 | 99.70 | 1.06 | 0.79 | 1.06 | 0.78 |
| 3 | VH837166 | 2 | 1,698 | 1,698 | 0.71 | 0.83 | 81.21 | 18.67 | 99.88 | 0.96 | 0.78 | 0.85 | 0.72 |
| 4 | VH840598 | 2 | 1,707 | 1,707 | 0.67 | 0.83 | 78.62 | 21.27 | 99.88 | 0.86 | 0.83 | 0.84 | 0.78 |
| 5 | VH858853 | 2 | 1,708 | 1,708 | 0.79 | 0.90 | 86.77 | 13.11 | 99.88 | 1.22 | 0.85 | 1.19 | 0.81 |
| 6 | VH875753 | 2 | 990 | 990 | 0.62 | 0.79 | 74.34 | 24.65 | 98.99 | 0.99 | 0.84 | 0.93 | 0.79 |
| 7 | VH882648 | 2 | 977 | 977 | 0.81 | 0.90 | 89.97 | 9.93 | 99.90 | 0.52 | 0.73 | 0.45 | 0.68 |
| 8 | VH882701 | 2 | 976 | 976 | 0.58 | 0.76 | 74.59 | 24.69 | 99.28 | 0.71 | 0.80 | 0.56 | 0.71 |
| 9 | VH887664 | 2 | 958 | 958 | 0.62 | 0.81 | 82.57 | 16.81 | 99.37 | 0.45 | 0.75 | 0.39 | 0.66 |
| **N/A** | **AVERAGE** | **N/A** | **1,299** | **1,299** | **0.69** | **0.81** | **81.44** | **18.21** | **99.65** | **0.79** | **0.76** | **0.75** | **0.71** |

Table 6.4 Interrater Reliability and Descriptive Statistics for the Ratings by AI and Human Raters in AI-Scoring of Operational Items for Grade Eight

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Prompt | Item ID | Score Points | Rater 1 N | Rater 2 N | Kappa | QWK | Percent Exact | Percent Adjacent | Percent Exact + Adjacent | Rater 1 Item Score Mean | Rater 1 Item Score SD | Rater 2 Item Score Mean | Rater 2 Item Score SD |
| 1 | VH728143 | 2 | 1,394 | 1,394 | 0.62 | 0.83 | 75.47 | 24.25 | 99.71 | 0.86 | 0.90 | 0.85 | 0.83 |
| 2 | VH803535 | 2 | 1,188 | 1,188 | 0.75 | 0.87 | 85.10 | 14.65 | 99.75 | 0.63 | 0.77 | 0.66 | 0.78 |
| 3 | VH811932 | 2 | 1,139 | 1,139 | 0.61 | 0.74 | 80.77 | 18.70 | 99.47 | 0.44 | 0.68 | 0.41 | 0.59 |
| 4 | VH834406 | 2 | 1,390 | 1,390 | 0.53 | 0.77 | 69.78 | 29.78 | 99.57 | 0.78 | 0.88 | 0.77 | 0.77 |
| 5 | VH851259 | 2 | 1,375 | 1,375 | 0.60 | 0.79 | 74.84 | 24.73 | 99.56 | 1.36 | 0.81 | 1.24 | 0.76 |
| 6 | VH858229 | 2 | 1,401 | 1,401 | 0.56 | 0.75 | 71.38 | 28.27 | 99.64 | 0.83 | 0.80 | 0.76 | 0.74 |
| 7 | VH858277 | 2 | 1,399 | 1,399 | 0.68 | 0.84 | 79.34 | 20.51 | 99.86 | 0.84 | 0.83 | 0.80 | 0.77 |
| 8 | VH882560 | 2 | 1,157 | 1,157 | 0.76 | 0.90 | 84.96 | 14.78 | 99.74 | 1.16 | 0.89 | 1.19 | 0.87 |
| 9 | VH882650 | 2 | 1,170 | 1,170 | 0.85 | 0.94 | 90.77 | 9.23 | 100.00 | 0.78 | 0.89 | 0.75 | 0.85 |
| **N/A** | **AVERAGE** | **N/A** | **1,290** | **1,290** | **0.66** | **0.83** | **79.16** | **20.54** | **99.70** | **0.85** | **0.83** | **0.83** | **0.77** |

Table 6.5 Interrater Reliability and Descriptive Statistics for the Ratings by AI and Human Raters in AI-Scoring of Operational Items for High School

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Prompt | Item ID | Score Points | Rater 1 N | Rater 2 N | Kappa | QWK | Percent Exact | Percent Adjacent | Percent Exact + Adjacent | Rater 1 Item Score Mean | Rater 1 Item Score SD | Rater 2 Item Score Mean | Rater 2 Item Score SD |
| 1 | VH651810 | 1 | 1,184 | 1,184 | 0.91 | 0.91 | 95.44 | 4.56 | 100.00 | 0.42 | 0.49 | 0.40 | 0.49 |
| 2 | VH651815 | 2 | 1,179 | 1,179 | 0.70 | 0.73 | 84.05 | 15.69 | 99.75 | 0.60 | 0.58 | 0.54 | 0.54 |
| 3 | VH835499 | 2 | 1,602 | 1,602 | 0.55 | 0.71 | 71.54 | 28.21 | 99.75 | 1.07 | 0.75 | 1.01 | 0.67 |
| 4 | VH849536 | 2 | 1,618 | 1,618 | 0.64 | 0.77 | 78.00 | 21.57 | 99.57 | 0.70 | 0.74 | 0.63 | 0.67 |
| 5 | VH852204 | 2 | 1,615 | 1,615 | 0.61 | 0.76 | 78.45 | 21.18 | 99.63 | 0.54 | 0.72 | 0.52 | 0.67 |
| 6 | VH853067 | 2 | 1,612 | 1,612 | 0.94 | 0.96 | 96.09 | 3.91 | 100.00 | 1.06 | 0.74 | 1.06 | 0.73 |
| 7 | VH870782 | 2 | 1,153 | 1,153 | 0.59 | 0.70 | 86.30 | 13.62 | 99.91 | 0.26 | 0.51 | 0.20 | 0.45 |
| 8 | VH876098 | 2 | 1,325 | 1,325 | 0.57 | 0.77 | 71.09 | 28.30 | 99.40 | 1.07 | 0.86 | 0.98 | 0.78 |
| 9 | VH876771 | 2 | 1,135 | 1,135 | 0.55 | 0.75 | 74.71 | 25.20 | 99.91 | 0.62 | 0.74 | 0.48 | 0.68 |
| **N/A** | **AVERAGE** | **N/A** | **1,380** | **1,380** | **0.67** | **0.78** | **81.74** | **18.03** | **99.77** | **0.70** | **0.68** | **0.65** | **0.63** |

Table 6.6 Interrater Reliability and Descriptive Statistics for the Ratings by Two Raters in Human-Scoring of Operational Items for Grade Eight

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Prompt | Item ID | Score Points | Rater 1 N | Rater 2 N | Kappa | QWK | Percent Exact | Percent Adjacent | Percent Exact + Adjacent | Rater 1 Item Score Mean | Rater 1 Item Score SD | Rater 2 Item Score Mean | Rater 2 Item Score SD |
| 1 | VH809632 | 1 | 1,165 | 1,165 | 0.62 | 0.62 | 86.61 | 13.39 | 100.00 | 0.22 | 0.42 | 0.24 | 0.43 |
| 2 | VH882698 | 2 | 1,170 | 1,170 | 0.64 | 0.78 | 79.40 | 20.34 | 99.74 | 0.57 | 0.70 | 0.56 | 0.69 |
| **N/A** | **AVERAGE** | **N/A** | **1,168** | **1,168** | **0.63** | **0.70** | **83.01** | **16.87** | **99.87** | **0.40** | **0.56** | **0.40** | **0.56** |

**Note:** Only grade eight had two human raters, for two items, in the 2020–2021 administration.

The 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 items were flagged if any of the following conditions occurred:

* Exact + adjacent agreement < 0.80
* Exact agreement < 0.70
* QWK < 0.70

Dichotomous items were flagged if either of the following conditions occurred:

* Exact agreement < 0.80
* QWK < 0.70

Table 6.7 shows the number of items flagged by grade level and the grade band, and by scoring method. There were three flagged items among the 29 operational items across grade levels. Of the three flagged items, one was flagged for human-human ratings and two for human-AI ratings. Flagged items were subsequently reviewed by content specialists.

Table 6.7 Number of Operational CR Items Flagged by Scoring Method

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Scoring Method | Grade Level | Flagged Polytomous Items | Flagged Dichotomous Items | Total Flagged Items | Total Number of Scored Items | Percent Flagged |
| AI scoring | Grade 5 | 0 | 1 | 1 | 9 | 11 |
| AI scoring | Grade 8 | 1 | 0 | 1 | 9 | 11 |
| AI scoring | High School | 0 | 0 | 0 | 9 | 0 |
| Human scoring | Grade 8 | 0 | 1 | 1 | 2 | 50 |

The evaluation of CR items with new, approved, AI models are presented in subsection [*6.4.3 Model Evaluation*](#_6.5.3_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 ALTRD 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 ALTRD 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.

#### Scoring for Selected-Response Items

CAST 2020–2021 assessments 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 a test 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 [*8.5* *Quality Control of Scoring*](#_Quality_Control_of_1).

### 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 the parallel scoring system developed by ETS Technology and Information Processing Services (TIPS). The ETS PAR team used this tool 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 process was an internal quality control step that is in place to verify the accuracy of scoring. Students’ scores were reported only when the two parallel systems produced identical results with acceptable tolerance.

Were scores not to match, the mismatch would be investigated by ETS’ PAR and eSKM teams and resolved. (For example, the mismatch could be a result of an ETS and CDE decision to not score an item as a problem was identified in a particular item or rubric.) ETS would apply a problem item notification (PIN) not to score the item through the systematic process in eSKM, which might result in a mismatch if TIPS were still in the process of applying the PIN in the parallel system when the student score was being compared. This real-time scoring check is designed to detect mismatches and track remediation.

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 2020–2021 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 [*7.4 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 test and domain scores.

Unlike the 2019–2020 administration, where CAASPP testing was suspended in spring 2020, resulting in small numbers of students having the opportunity to take the CAST, testing was fully available in 2020–2021. This permitted typical psychometric analyses of the CAST scores to be conducted and presented in this technical report. These results include reporting individual student theta score distributions, which are presented in table 6.A.1 through table 6.A.6 in [appendix 6.A](#_Appendix_6.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

The CAST uses an IRT model to estimate students’ abilities (i.e., 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 comprised of 100 Rasch items with a difficulty of 0. A total theta score is converted to an NR score using equation 6.1. *Refer to the* [*Alternative Text for Equation 6.1*](#_Alternative_Text_for_7) *for a description of this equation.*

 (6.1)

where,

*n* = 100.

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 6.8 shows the scaling constants for the linear transformation of an NR score to a scale score.

Table 6.8 Scaling Constants

|  |  |  |
| --- | --- | --- |
| Grade Level | 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 6.8.

The ranges of the reporting scale scores are 150–250, 350–450, and 550–650 for grade five, grade eight, and high school, respectively.

The CAST reports scale scores for the total test for students who have answered at least 10 items. Those students who did not answer any items on the test 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 at least a minimum number of operational items for the total test. Table 6.9 lists the minimum number of operational items required to fulfill the completion requirement for the CAST for the 2020–‍2021 administration. For example, the minimum number of items required for grade five is 45.

Table 6.9 Minimum Number of Item Requirements for Test Completion

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grade Level | Minimum Number of Items | Life Sciences | Physical Sciences | Earth and Space Sciences |
| Grade 5 | 45 | 8 | 8 | 7 |
| Grade 8 | 45 | 8 | 8 | 8 |
| High school | 44 | 8 | 8 | 8 |

For students who answered at least 10 items but did not complete the test, 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 test completed.

Table 6.10 shows the mean and SD of both scale scores and theta scores for the CAST.

Table 6.10 Mean and SD of Theta Scores and Scale Scores

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grade Level | Number of Students | Scale Score Mean | Scale Score SD | Theta Score Mean | Theta Score SD |
| Grade 5 | 62,238 | 199 | 22.3 | -0.07 | 1.07 |
| Grade 8 | 75,000 | 399 | 22.8 | -0.04 | 1.09 |
| High school—Grade 10 | 2,270 | 605 | 22.5 | 0.26 | 1.13 |
| High school—Grade 11 | 43,791 | 602 | 22.2 | 0.09 | 1.07 |
| High school—Grade 12 | 61,723 | 600 | 21.4 | -0.01 | 1.03 |
| High school—All grades | 107,784 | 601 | 21.8 | 0.04 | 1.05 |

Individual student scale score distributions are presented in table 6.B.1 through table 6.B.6 in [appendix 6.B](#_Appendix_6.B:_Overall) for grades five, eight, ten, eleven, and twelve, as well as for the high school grade band.

#### Achievement Levels

A standard setting was conducted after the 2018–2019 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 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 performance expectations 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 performance expectations 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 performance expectations of the CA NGSS.

The scale score ranges for achievement levels are shown in table 6.11. Percentages of students in each achievement level are in table 6.12, and their graphic representation is displayed in figure 6.1.

Table 6.11 Scale Score Ranges for Achievement Levels

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grade Level | 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 |

Table 6.12 Percent of Students in Each Achievement Level for Total Scores

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | 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 | 62,238 | 21.93 | 48.32 | 19.80 | 9.94 | 29.75 |
| Grade 8 | 75,000 | 21.02 | 51.96 | 18.10 | 8.92 | 27.02 |
| High school—Grade 10 | 2,270 | 10.40 | 54.80 | 23.22 | 11.59 | 34.80 |
| High school—Grade 11 | 43,791 | 13.87 | 54.63 | 24.27 | 7.22 | 31.49 |
| High school—Grade 12 | 61,723 | 14.32 | 58.14 | 21.84 | 5.70 | 27.54 |
| High school—All grades | 107,784 | 14.06 | 56.64 | 22.85 | 6.44 | 29.30 |

Figure 6.1 Percentage of achievement levels

Overall, 27 to 30 percent of students in grades five and eight, and in the high school grade band met or exceeded the standards. For high school students, the percentage was the lowest for grade twelve and the highest for grade ten.

Demographic summaries of achievement levels for total scores are presented in table 6.C.1 through table 6.C.6 in [appendix 6.C](#_Appendix_6.C:_Demographic) for grades five, eight, ten, eleven, and twelve, as well as for the high school grade band. The description of the demographic student groups included in these tables are presented in table 5.2 in [*Chapter 5: Test Administration*](#_Chapter_5:_Test). The percentages of students who met or exceeded standards were similar between male and female students.

Those student groups with high percentages of students who met or exceeded standards were initial fluent English proficient (IFEP), Asian, or not economically disadvantaged for grades five and eight and for high school. Additionally, for high school, Filipino and nonbinary student groups were among the highest performing. In contrast, those groups with low percentages of students who met or exceeded standards for all grade levels were the English learner (EL) and migrant education student groups. Included in the lowest performing groups for grade eight and high school were the special education, homeless (for grade eight), and adult EL (for high school) student groups.

#### Domain Achievement Levels

Science domain achievement levels are reported for students who have met the minimum number of item requirements for the science domain as listed in table 6.9. Students might receive science domain achievement levels for some domain(s) but not the others depending on the number of items they completed for different domains. Science domain achievement levels are not reported for students who received the LOSS (i.e., answered fewer than 10 items for the total test).

There are three domains for each test: Life Sciences, Physical Sciences, and Earth and Space Sciences. The detailed descriptions of the three achievement levels for each science domain are shown in table 6.13.

Table 6.13 Description of Science Domain Achievement Levels

|  |  |  |  |
| --- | --- | --- | --- |
| Science Domain | Below Standard | Near Standard | Above Standard |
| 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. |

Table 6.13 *(continuation)*

|  |  |  |  |
| --- | --- | --- | --- |
| 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. |

A student is assigned to one of the three achievement levels for a science domain according to the following rules:

* Place in the Below Standard level if *θd* < *θ–3*− 1.5
* Place in the Near Standard level if *θ–3*− 1.5 ≤ *θd* < *θL3*+ 1.5
* Place in the Above Standard level if *θd* ≥ *θL3*+ 1.5

where

*θd* is a science domain theta score,

is the mean standard error of the science domain theta scores, and

*θL3* is the level 3 theta threshold score of the total test.

This science domain achievement level estimation method for CAST is similar to, but slightly different from, the Smarter Balanced approach for a claim performance level. The CAST uses the mean standard error of science domain theta scores, while Smarter Balanced uses the standard error of an individual student claim theta score.

Table 6.14 through table 6.16 show the percentages of science domain achievement levels for Life Sciences, Physical Sciences, and Earth and Space Sciences, respectively.

Table 6.14 Percent of Students in Each Achievement Level for the Life Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grade Level | Number of Students Tested | Below Standard | Near Standard | Above Standard |
| Grade 5 | 61,684 | 44.72 | 44.14 | 11.13 |
| Grade 8 | 72,308 | 40.69 | 50.49 | 8.82 |
| High school—Grade 10 | 2,229 | 28.13 | 56.62 | 15.25 |
| High school—Grade 11 | 42,955 | 32.03 | 57.69 | 10.28 |
| High school—Grade 12 | 60,651 | 34.62 | 57.00 | 8.38 |
| High school—All grades | 105,835 | 33.44 | 57.27 | 9.29 |

Table 6.15 Percent of Students in Each Achievement Level for the Physical Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grade Level | Number of Students Tested | Below Standard | Near Standard | Above Standard |
| Grade 5 | 61,763 | 47.74 | 40.84 | 11.42 |
| Grade 8 | 72,654 | 40.12 | 50.18 | 9.71 |
| High school—Grade 10 | 2,232 | 36.07 | 51.39 | 12.54 |
| High school—Grade 11 | 42,982 | 40.85 | 49.90 | 9.26 |
| High school—Grade 12 | 60,701 | 43.65 | 48.56 | 7.79 |
| High school—All grades | 105,915 | 42.35 | 49.16 | 8.49 |

Table 6.16 Percent of Students in Each Achievement Level for the Earth and Space Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grade Level | Number of Students Tested | Below Standard | Near Standard | Above Standard |
| Grade 5 | 61,607 | 42.13 | 46.60 | 11.28 |
| Grade 8 | 72,407 | 40.84 | 47.00 | 12.16 |
| High school—Grade 10 | 2,233 | 33.90 | 48.90 | 17.20 |
| High school—Grade 11 | 43,005 | 37.31 | 49.05 | 13.64 |
| High school—Grade 12 | 60,746 | 40.33 | 48.53 | 11.14 |
| High school—All grades | 105,984 | 38.97 | 48.75 | 12.28 |

Across all grade levels, the percentages of students in the Above Standard achievement level ranged from 8.4 to 15.3 for Life Sciences, from 7.8 to 12.5 for Physical Sciences, and from 11.1 to 17.2 for Earth and Space Sciences. For all grade levels across science domains, 52.3 to 71.9 percent of students achieved either Near Standard or Above Standard levels.

The percentages of students identified as Below Standard were lower than the percentages of Near Standard for all grade levels and science domains with the exception of grade five, for which the Life Sciences and Physical Sciences percentages were the same or higher for students identified as Below Standard.

Demographic summaries of science domain achievement levels are presented in [appendix 6.D](#_Appendix_6.D:_Demographic) for grades five, eight, ten, eleven, and twelve, and for high school. Table 6.D.1 through table 6.D.6 are for Life Sciences, table 6.D.7 through table 6.D.12 are for Physical Sciences, and table 6.D.13 through table 6.D.18 are for Earth and Space Sciences. The description of the demographic student groups included in these tables is shown in table 5.2 in [*Chapter 5: Test Administration*](#_Chapter_5:_Test).

#### Theta Scores Standard Errors

The conditional standard error of measurement (CSEM) is the SD of the distribution of theta scores that the student would earn under different testing conditions.

In the framework of IRT, the CSEM is the reciprocal of the square root of the test information function (TIF) based on the items taken by each student. It is also the estimate of standard error for the estimate of theta. The TIF is the sum of information from each item on the test. The CSEM for a student with proficiency *θj* is presented in equation 6.2. *Refer to the* [*Alternative Text for Equation 6.2*](#_Alternative_Text_for_8) *for a description of this equation.*

 (6.2)

where,

*I(θj)* is the test information for student *j*, calculated as presented in equation 6.3 *(Refer to the* [*Alternative Text for Equation 6.3*](#_Alternative_Text_for_9) *for a description of this equation.)*

 (6.3)

and *Ii*(*θj*)is the item information of item *i* for student *j*.

Item information based on the generalized partial credit model for both dichotomous and polytomous items is calculated as presented in equation 6.4. *Refer to the* [*Alternative Text for Equation 6.4*](#_Alternative_Text_for_10) *for a description of this equation.*

 (6.4)

where,

*Si*(*θj*) is the expected item score for item *i* on a theta score *θj*, calculated as presented in equation 6.5 (*Refer to the* [*Alternative Text for Equation 6.5*](#_Alternative_Text_for_11) *for a description of this equation.*)

 (6.5)

and equation 6.6 (*Refer to the* [*Alternative Text for Equation 6.6*](#_Alternative_Text_for_12) *for a description of this equation.*)

 (6.6)

where,

*Pih*(*θj*) is the probability of an examinee with *θj* getting score *h* on item *i*, the computation of which is shown in equation 7.7,

*ni* is the maximum number of score points for item *i*, and

*D* is a scaling constant of 1.7 that makes the logistic model approximate the normal ogive model.

The CSEM is calculated based only on the answered item(s) for both complete and incomplete tests.

#### Scale Scores Standard Errors

The conditional standard errors of theta scores can be transformed onto the reporting scale. This transformation is presented in equation 6.7. *Refer to the* [*Alternative Text for Equation 6.7*](#_Alternative_Text_for_13) *for a description of this equation.*

 (6.7)

where,

*n*=100,

*A* is the scaling constant that equals to 1.0081 as defined in table 6.8, and

SE sub theta sub jis the standard error estimate of *θj* defined in equation 6.2.

### Reports Produced and Scores for Each Report

The CAST provides 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 CAASPP reporting for students. Scores for the CAST, as one of the components in the CAASPP, are reported through the CAASPP reporting system.

#### Online Reporting

TOMS is a secure website hosted by ETS that permits LEA users to manage the CAASPP computer-based summative 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 (i.e., 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 2020–2021 administration: the California Educator Reporting System (CERS).

TOMS communicated with CERS, which provided authorized users with interactive and cumulative online reports for the CAST at the student, school, and LEA levels. CERS provided preliminary score data for each administered test available in the reporting system.

Based on the CAST reporting requirements, CERS provided the preliminary summative reports containing information outlining student knowledge and skills, as well as performance levels aligned with the assessment-specific performance areas. CERS also permitted access to individual score reports, which provided preliminary score data for each administered test available in the reporting system. The online aggregated reports were available to be downloaded in PDF, Excel, and comma-separated value formats.

CERS was the primary source for LEA staff to analyze CAST results at the LEA, school, grade, classroom, or customized group level. CERS provided these reports, which can be downloaded to plan instruction. LEA staff with TOMS logon credentials could enter CERS through the CAASPP 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 two categories of CAST reports. The specific reports within each category are presented in this subsection.

* **Student Score Report (SSR)—**The SSR was the official score report for parents and guardians. An SSR described the student’s results.
* **LEA student data files and aggregations—**LEA student data files were available for download on demand by the LEA in TOMS to coincide with availability of the SSRs.

##### 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 6.11.)
* Reported achievement levels (CAST achievement levels are “Standard Exceeded,” “Standard Met,” “Standard Nearly Met,” and “Standard Not Met.”)
* Reported science-domain-specific achievement levels (The science-domain-specific achievement levels are “Above Standard,” “Near Standard,” and “Below Standard.”)

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 [*2.4.1 Universal Tools, Designated Supports, and Accommodations*](#_Toc93326451).

LEAs had three options for accessing and distributing SSRs to parents/guardians:

1. Accessing electronic SSR PDFs using a locally provided parent or student portal
2. Downloading SSR PDFs 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 forward 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 or guardian. CAST SSRs that included individual student results were not distributed beyond the student’s school.

Further information about the SSR and its interpretation is provided on the Smarter Balanced 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 forward the appropriate reports to test sites. Optionally, the LEA could download and then print the SSR PDF and then send 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 CAASPP test site coordinator to download from TOMS.

Preliminary 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 tests taken.

Current and historical aggregated results are accessible to the public on the CDE Test Results for California’s Assessments website.

#### Score Report Applications

CAST results provide parents/guardians with information about their child’s progress. The results are 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 may 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 help schools and LEAs identify strengths and weaknesses 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 CAST results to help make decisions about student placement, promotion, retention, or other considerations related to student achievement. However, it is important to remember that a single test can provide only limited information. Other relevant information should be considered as well. It is advisable for parents/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 CAST computer-based assessment results. It is also important to note that a student’s score 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 for high school. 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 CAST user can also make comparisons within the same grade level or grade band across years.

However, comparing scale scores from different grades for the CAST 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 Smarter Balanced Starting Smarter website for California assessments or the *CAASPP Post-Test Guide* (CDE, 2021).

### New Artificial Intelligence Model Building

ETS built models for 13 field test items and two operational items that were human scored during the 2020–2021 administration. The breakdown of item counts by grade level is shown in table 6.17:

Table 6.17 Number of Items for New AI Model Building by Grade Level

|  |  |  |  |
| --- | --- | --- | --- |
| Grade Level | Number of Field Test Items | Number of Operational Items | Total |
| Grade 5 | 0 | 0 | 0 |
| Grade 8 | 0 | 2 | 2 |
| High school | 13 | 0 | 13 |
| **Total** | **13** | **2** | **15** |

Of the 15 AI models that were built, 13 were approved. The evaluation process of the AI models is presented in subsection [*6.4.3 Model Evaluation*](#_6.5.3_Model_Evaluation).

#### Data Collection

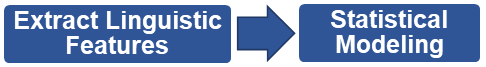
After the CAST administration, ETS collected a sample of students’ responses to 15 CR items with human score(s) assigned, as described in subsection [*6.1.1.1 Sampling Process*](#_Sampling_Process).

#### 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).[[11]](#footnote-12) Each model was built using a 10-fold cross-validation method that randomly split the entire dataset for an item in 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 6.2 provides a cycle chart illustrating the primary steps in the model-building and evaluation processes. First, three human-scored responses with scores of 1, 1, and 2 are funneled to natural language processing tools to extract linguistic features. An arrow points to the next step, statistical modeling. Here, the model-building process ends. The resulting model from the previous steps is sent to model evaluation.



**Model Evaluation**

**Model Building**

**Scored Responses**

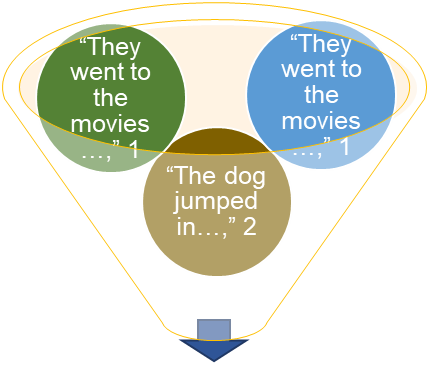


Figure 6.2 Model building and evaluation process

#### 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 large enough sample of responses and evaluate the extent to which they agree with each other.

Each item response had 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, human–AI agreement, and the comparison of the two types of agreement. 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 as expected.

### References

California Department of Education. (2021). CAASPP post-test guide: Technical information for student score reports for CAASPP LEA and test site coordinators and research specialists. 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. *Advances in Neural Information Processing Systems 9*, 155–161. 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 6.1

NR is equal to n times the fraction with the numerator exponent based theta sub j and the denominator one plus exponent theta sub j.

#### Alternative Text for Equation 6.2

SEM of Theta sub j equals 1 divided by the square root of I of theta sub j.

#### Alternative Text for Equation 6.3

I of Theta sub j equals the sum from i equals 1 to n of I sub i of Theta sub j.

#### Alternative Text for Equation 6.4

I sub i of Theta sub j equals open parenthesis D times a sub i close parenthesis squared times open bracket s sub i2 of theta sub j minus s squared sub i of theta sub j close bracket.

#### Alternative Text for Equation 6.5

S sub i of Theta sub j equals the sum from h equals 0 to n sub i of h times p sub ih of theta sub j.

#### Alternative Text for Equation 6.6

S sub i2 of Theta sub j equals the sum from h equals 0 to n sub i of h squared times p sub ih of theta sub j.

#### Alternative Text for Equation 6.7

CSEM of sub scale score equals A times SE of theta sub j times n times the fraction with the numerator exponent based theta sub j and the denominator open parenthesis of one plus exponent theta sub j close parenthesis square.

### Appendix 6.A: Overall Theta Score Distribution

Table 6.A.1 Overall Theta Score Distribution for Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Theta Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| θ = -4.0 | 4 | 0.0 | 4 | 0.0 |
| -4.0 < θ ≤ -3.5 | 11 | 0.0 | 15 | 0.0 |
| -3.5 < θ ≤ -3.0 | 31 | 0.0 | 46 | 0.1 |
| -3.0 < θ ≤ -2.5 | 230 | 0.4 | 276 | 0.4 |
| -2.5 < θ ≤ -2.0 | 1,178 | 1.9 | 1,454 | 2.3 |
| -2.0 < θ ≤ -1.5 | 3,763 | 6.0 | 5,217 | 8.4 |
| -1.5 < θ ≤ -1.0 | 7,367 | 11.8 | 12,584 | 20.2 |
| -1.0 < θ ≤ -0.5 | 9,900 | 15.9 | 22,484 | 36.1 |
| -0.5 < θ ≤ 0 | 10,848 | 17.4 | 33,332 | 53.6 |
| 0 < θ ≤ 0.5 | 10,671 | 17.1 | 44,003 | 70.7 |
| 0.5 < θ ≤ 1.0 | 8,427 | 13.5 | 52,430 | 84.2 |
| 1.0 < θ ≤ 1.5 | 5,268 | 8.5 | 57,698 | 92.7 |
| 1.5 < θ ≤ 2.0 | 2,644 | 4.2 | 60,342 | 97.0 |
| 2.0 < θ ≤ 2.5 | 1,068 | 1.7 | 61,410 | 98.7 |
| 2.5 < θ ≤ 3.0 | 546 | 0.9 | 61,956 | 99.5 |
| 3.0 < θ ≤ 3.5 | 164 | 0.3 | 62,120 | 99.8 |
| 3.5 < θ < 4.0 | 60 | 0.1 | 62,180 | 99.9 |
| θ = 4.0 | 58 | 0.1 | 62,238 | 100.0 |

Table 6.A.2 Overall Theta Score Distribution for Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Theta Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| θ = -4.0 | 28 | 0.0 | 28 | 0.0 |
| -4.0 < θ ≤ -3.5 | 19 | 0.0 | 47 | 0.1 |
| -3.5 < θ ≤ -3.0 | 81 | 0.1 | 128 | 0.2 |
| -3.0 < θ ≤ -2.5 | 245 | 0.3 | 373 | 0.5 |
| -2.5 < θ ≤ -2.0 | 1,354 | 1.8 | 1,727 | 2.3 |
| -2.0 < θ ≤ -1.5 | 3,956 | 5.3 | 5,683 | 7.6 |
| -1.5 < θ ≤ -1.0 | 8,252 | 11.0 | 13,935 | 18.6 |
| -1.0 < θ ≤ -0.5 | 12,599 | 16.8 | 26,534 | 35.4 |
| -0.5 < θ ≤ 0 | 13,768 | 18.4 | 40,302 | 53.7 |
| 0 < θ ≤ 0.5 | 12,814 | 17.1 | 53,116 | 70.8 |
| 0.5 < θ ≤ 1.0 | 9,671 | 12.9 | 62,787 | 83.7 |
| 1.0 < θ ≤ 1.5 | 5,947 | 7.9 | 68,734 | 91.6 |
| 1.5 < θ ≤ 2.0 | 3,244 | 4.3 | 71,978 | 96.0 |
| 2.0 < θ ≤ 2.5 | 1,587 | 2.1 | 73,565 | 98.1 |
| 2.5 < θ ≤ 3.0 | 850 | 1.1 | 74,415 | 99.2 |
| 3.0 < θ ≤ 3.5 | 324 | 0.4 | 74,739 | 99.7 |
| 3.5 < θ < 4.0 | 142 | 0.2 | 74,881 | 99.8 |
| θ = 4.0 | 119 | 0.2 | 75,000 | 100.0 |

Table 6.A.3 Overall Theta Score Distribution for Grade Ten

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Theta Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| θ = -4.0 | 0 | 0.0 | 0 | 0.0 |
| -4.0 < θ ≤ -3.5 | 1 | 0.0 | 1 | 0.0 |
| -3.5 < θ ≤ -3.0 | 0 | 0.0 | 1 | 0.0 |
| -3.0 < θ ≤ -2.5 | 3 | 0.1 | 4 | 0.2 |
| -2.5 < θ ≤ -2.0 | 13 | 0.6 | 17 | 0.7 |
| -2.0 < θ ≤ -1.5 | 82 | 3.6 | 99 | 4.4 |
| -1.5 < θ ≤ -1.0 | 178 | 7.8 | 277 | 12.2 |
| -1.0 < θ ≤ -0.5 | 321 | 14.1 | 598 | 26.3 |
| -0.5 < θ ≤ 0 | 392 | 17.3 | 990 | 43.6 |
| 0 < θ ≤ 0.5 | 412 | 18.1 | 1,402 | 61.8 |
| 0.5 < θ ≤ 1.0 | 347 | 15.3 | 1,749 | 77.0 |
| 1.0 < θ ≤ 1.5 | 202 | 8.9 | 1,951 | 85.9 |
| 1.5 < θ ≤ 2.0 | 139 | 6.1 | 2,090 | 92.1 |
| 2.0 < θ ≤ 2.5 | 86 | 3.8 | 2,176 | 95.9 |
| 2.5 < θ ≤ 3.0 | 62 | 2.7 | 2,238 | 98.6 |
| 3.0 < θ ≤ 3.5 | 16 | 0.7 | 2,254 | 99.3 |
| 3.5 < θ < 4.0 | 6 | 0.3 | 2,260 | 99.6 |
| θ = 4.0 | 10 | 0.4 | 2,270 | 100.0 |

Table 6.A.4 Overall Theta Score Distribution for Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Theta Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| θ = -4.0 | 6 | 0.0 | 6 | 0.0 |
| -4.0 < θ ≤ -3.5 | 14 | 0.0 | 20 | 0.0 |
| -3.5 < θ ≤ -3.0 | 14 | 0.0 | 34 | 0.1 |
| -3.0 < θ ≤ -2.5 | 125 | 0.3 | 159 | 0.4 |
| -2.5 < θ ≤ -2.0 | 544 | 1.2 | 703 | 1.6 |
| -2.0 < θ ≤ -1.5 | 1,848 | 4.2 | 2,551 | 5.8 |
| -1.5 < θ ≤ -1.0 | 4,236 | 9.7 | 6,787 | 15.5 |
| -1.0 < θ ≤ -0.5 | 6,403 | 14.6 | 13,190 | 30.1 |
| -0.5 < θ ≤ 0 | 7,801 | 17.8 | 20,991 | 47.9 |
| 0 < θ ≤ 0.5 | 7,849 | 17.9 | 28,840 | 65.9 |
| 0.5 < θ ≤ 1.0 | 6,790 | 15.5 | 35,630 | 81.4 |
| 1.0 < θ ≤ 1.5 | 4,070 | 9.3 | 39,700 | 90.7 |
| 1.5 < θ ≤ 2.0 | 2,266 | 5.2 | 41,966 | 95.8 |
| 2.0 < θ ≤ 2.5 | 967 | 2.2 | 42,933 | 98.0 |
| 2.5 < θ ≤ 3.0 | 532 | 1.2 | 43,465 | 99.3 |
| 3.0 < θ ≤ 3.5 | 183 | 0.4 | 43,648 | 99.7 |
| 3.5 < θ < 4.0 | 80 | 0.2 | 43,728 | 99.9 |
| θ = 4.0 | 63 | 0.1 | 43,791 | 100.0 |

Table 6.A.5 Overall Theta Score Distribution for Grade Twelve

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Theta Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| θ = -4.0 | 17 | 0.0 | 17 | 0.0 |
| -4.0 < θ ≤ -3.5 | 16 | 0.0 | 33 | 0.1 |
| -3.5 < θ ≤ -3.0 | 13 | 0.0 | 46 | 0.1 |
| -3.0 < θ ≤ -2.5 | 194 | 0.3 | 240 | 0.4 |
| -2.5 < θ ≤ -2.0 | 754 | 1.2 | 994 | 1.6 |
| -2.0 < θ ≤ -1.5 | 2,668 | 4.3 | 3,662 | 5.9 |
| -1.5 < θ ≤ -1.0 | 6,483 | 10.5 | 10,145 | 16.4 |
| -1.0 < θ ≤ -0.5 | 9,874 | 16.0 | 20,019 | 32.4 |
| -0.5 < θ ≤ 0 | 11,944 | 19.4 | 31,963 | 51.8 |
| 0 < θ ≤ 0.5 | 11,179 | 18.1 | 43,142 | 69.9 |
| 0.5 < θ ≤ 1.0 | 9,069 | 14.7 | 52,211 | 84.6 |
| 1.0 < θ ≤ 1.5 | 4,929 | 8.0 | 57,140 | 92.6 |
| 1.5 < θ ≤ 2.0 | 2,548 | 4.1 | 59,688 | 96.7 |
| 2.0 < θ ≤ 2.5 | 1,109 | 1.8 | 60,797 | 98.5 |
| 2.5 < θ ≤ 3.0 | 559 | 0.9 | 61,356 | 99.4 |
| 3.0 < θ ≤ 3.5 | 210 | 0.3 | 61,566 | 99.7 |
| 3.5 < θ < 4.0 | 85 | 0.1 | 61,651 | 99.9 |
| θ = 4.0 | 72 | 0.1 | 61,723 | 100.0 |

Table 6.A.6 Overall Theta Score Distribution for High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Theta Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| θ = -4.0 | 23 | 0.0 | 23 | 0.0 |
| -4.0 < θ ≤ -3.5 | 31 | 0.0 | 54 | 0.1 |
| -3.5 < θ ≤ -3.0 | 27 | 0.0 | 81 | 0.1 |
| -3.0 < θ ≤ -2.5 | 322 | 0.3 | 403 | 0.4 |
| -2.5 < θ ≤ -2.0 | 1,311 | 1.2 | 1,714 | 1.6 |
| -2.0 < θ ≤ -1.5 | 4,598 | 4.3 | 6,312 | 5.9 |
| -1.5 < θ ≤ -1.0 | 10,897 | 10.1 | 17,209 | 16.0 |
| -1.0 < θ ≤ -0.5 | 16,598 | 15.4 | 33,807 | 31.4 |
| -0.5 < θ ≤ 0 | 20,137 | 18.7 | 53,944 | 50.0 |
| 0 < θ ≤ 0.5 | 19,440 | 18.0 | 73,384 | 68.1 |
| 0.5 < θ ≤ 1.0 | 16,206 | 15.0 | 89,590 | 83.1 |
| 1.0 < θ ≤ 1.5 | 9,201 | 8.5 | 98,791 | 91.7 |
| 1.5 < θ ≤ 2.0 | 4,953 | 4.6 | 103,744 | 96.3 |
| 2.0 < θ ≤ 2.5 | 2,162 | 2.0 | 105,906 | 98.3 |
| 2.5 < θ ≤ 3.0 | 1,153 | 1.1 | 107,059 | 99.3 |
| 3.0 < θ ≤ 3.5 | 409 | 0.4 | 107,468 | 99.7 |
| 3.5 < θ < 4.0 | 171 | 0.2 | 107,639 | 99.9 |
| θ = 4.0 | 145 | 0.1 | 107,784 | 100.0 |

### Appendix 6.B: Overall Scale Score Distribution

Table 6.B.1 Overall Scale Score Distribution for Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 150 | 264 | 0.4 | 264 | 0.4 |
| 151 | 1 | 0.0 | 265 | 0.4 |
| 152 | 8 | 0.0 | 273 | 0.4 |
| 153 | 12 | 0.0 | 285 | 0.5 |
| 154 | 30 | 0.0 | 315 | 0.5 |
| 155 | 43 | 0.1 | 358 | 0.6 |
| 156 | 43 | 0.1 | 401 | 0.6 |
| 157 | 77 | 0.1 | 478 | 0.8 |
| 158 | 165 | 0.3 | 643 | 1.0 |
| 159 | 198 | 0.3 | 841 | 1.3 |
| 160 | 176 | 0.3 | 1,017 | 1.6 |
| 161 | 379 | 0.6 | 1,396 | 2.2 |
| 162 | 363 | 0.6 | 1,759 | 2.8 |
| 163 | 388 | 0.6 | 2,147 | 3.4 |
| 164 | 576 | 0.9 | 2,723 | 4.4 |
| 165 | 433 | 0.7 | 3,156 | 5.0 |
| 166 | 541 | 0.9 | 3,697 | 5.9 |
| 167 | 840 | 1.3 | 4,537 | 7.3 |
| 168 | 784 | 1.3 | 5,321 | 8.5 |
| 169 | 727 | 1.2 | 6,048 | 9.7 |
| 170 | 788 | 1.3 | 6,836 | 10.9 |
| 171 | 767 | 1.2 | 7,603 | 12.2 |
| 172 | 967 | 1.5 | 8,570 | 13.7 |
| 173 | 830 | 1.3 | 9,400 | 15.0 |
| 174 | 836 | 1.3 | 10,236 | 16.4 |
| 175 | 880 | 1.4 | 11,116 | 17.8 |
| 176 | 1,004 | 1.6 | 12,120 | 19.4 |
| 177 | 830 | 1.3 | 12,950 | 20.7 |
| 178 | 963 | 1.5 | 13,913 | 22.3 |
| 179 | 865 | 1.4 | 14,778 | 23.6 |
| 180 | 649 | 1.0 | 15,427 | 24.7 |
| 181 | 1,155 | 1.8 | 16,582 | 26.5 |
| 182 | 900 | 1.4 | 17,482 | 28.0 |
| 183 | 947 | 1.5 | 18,429 | 29.5 |
| 184 | 1,005 | 1.6 | 19,434 | 31.1 |
| 185 | 649 | 1.0 | 20,083 | 32.1 |
| 186 | 979 | 1.6 | 21,062 | 33.7 |
| 187 | 868 | 1.4 | 21,930 | 35.1 |

Table 6.B.1 *(continuation one)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 188 | 913 | 1.5 | 22,843 | 36.5 |
| 189 | 1,083 | 1.7 | 23,926 | 38.3 |
| 190 | 512 | 0.8 | 24,438 | 39.1 |
| 191 | 1,130 | 1.8 | 25,568 | 40.9 |
| 192 | 766 | 1.2 | 26,334 | 42.1 |
| 193 | 886 | 1.4 | 27,220 | 43.6 |
| 194 | 875 | 1.4 | 28,095 | 45.0 |
| 195 | 864 | 1.4 | 28,959 | 46.3 |
| 196 | 1,044 | 1.7 | 30,003 | 48.0 |
| 197 | 591 | 0.9 | 30,594 | 48.9 |
| 198 | 1,017 | 1.6 | 31,611 | 50.6 |
| 199 | 552 | 0.9 | 32,163 | 51.5 |
| 200 | 1,043 | 1.7 | 33,206 | 53.1 |
| 201 | 774 | 1.2 | 33,980 | 54.4 |
| 202 | 935 | 1.5 | 34,915 | 55.9 |
| 203 | 687 | 1.1 | 35,602 | 57.0 |
| 204 | 1,317 | 2.1 | 36,919 | 59.1 |
| 205 | 626 | 1.0 | 37,545 | 60.1 |
| 206 | 935 | 1.5 | 38,480 | 61.6 |
| 207 | 677 | 1.1 | 39,157 | 62.6 |
| 208 | 944 | 1.5 | 40,101 | 64.2 |
| 209 | 667 | 1.1 | 40,768 | 65.2 |
| 210 | 951 | 1.5 | 41,719 | 66.7 |
| 211 | 668 | 1.1 | 42,387 | 67.8 |
| 212 | 947 | 1.5 | 43,334 | 69.3 |
| 213 | 654 | 1.0 | 43,988 | 70.4 |
| 214 | 976 | 1.6 | 44,964 | 71.9 |
| 215 | 686 | 1.1 | 45,650 | 73.0 |
| 216 | 834 | 1.3 | 46,484 | 74.4 |
| 217 | 557 | 0.9 | 47,041 | 75.3 |
| 218 | 1,150 | 1.8 | 48,191 | 77.1 |
| 219 | 534 | 0.9 | 48,725 | 78.0 |
| 220 | 987 | 1.6 | 49,712 | 79.5 |
| 221 | 600 | 1.0 | 50,312 | 80.5 |
| 222 | 865 | 1.4 | 51,177 | 81.9 |
| 223 | 526 | 0.8 | 51,703 | 82.7 |
| 224 | 903 | 1.4 | 52,606 | 84.2 |
| 225 | 510 | 0.8 | 53,116 | 85.0 |

Table 6.B.1 *(continuation two)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 226 | 748 | 1.2 | 53,864 | 86.2 |
| 227 | 573 | 0.9 | 54,437 | 87.1 |
| 228 | 746 | 1.2 | 55,183 | 88.3 |
| 229 | 675 | 1.1 | 55,858 | 89.4 |
| 230 | 456 | 0.7 | 56,314 | 90.1 |
| 231 | 706 | 1.1 | 57,020 | 91.2 |
| 232 | 305 | 0.5 | 57,325 | 91.7 |
| 233 | 748 | 1.2 | 58,073 | 92.9 |
| 234 | 361 | 0.6 | 58,434 | 93.5 |
| 235 | 529 | 0.8 | 58,963 | 94.3 |
| 236 | 328 | 0.5 | 59,291 | 94.9 |
| 237 | 541 | 0.9 | 59,832 | 95.7 |
| 238 | 197 | 0.3 | 60,029 | 96.0 |
| 239 | 275 | 0.4 | 60,304 | 96.5 |
| 240 | 499 | 0.8 | 60,803 | 97.3 |
| 241 | 218 | 0.3 | 61,021 | 97.6 |
| 242 | 382 | 0.6 | 61,403 | 98.2 |
| 243 | 153 | 0.2 | 61,556 | 98.5 |
| 244 | 313 | 0.5 | 61,869 | 99.0 |
| 245 | 88 | 0.1 | 61,957 | 99.1 |
| 246 | 214 | 0.3 | 62,171 | 99.5 |
| 247 | 74 | 0.1 | 62,245 | 99.6 |
| 248 | 120 | 0.2 | 62,365 | 99.8 |
| 249 | 28 | 0.0 | 62,393 | 99.8 |
| 250 | 109 | 0.2 | 62,502 | 100.0 |

Table 6.B.2 Overall Scale Score Distribution for Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 350 | 1,370 | 1.8 | 1,370 | 1.8 |
| 351 | 16 | 0.0 | 1,386 | 1.8 |
| 352 | 16 | 0.0 | 1,402 | 1.8 |
| 353 | 43 | 0.1 | 1,445 | 1.9 |
| 354 | 60 | 0.1 | 1,505 | 2.0 |
| 355 | 148 | 0.2 | 1,653 | 2.2 |
| 356 | 82 | 0.1 | 1,735 | 2.3 |
| 357 | 191 | 0.3 | 1,926 | 2.5 |
| 358 | 299 | 0.4 | 2,225 | 2.9 |
| 359 | 201 | 0.3 | 2,426 | 3.2 |
| 360 | 509 | 0.7 | 2,935 | 3.8 |
| 361 | 322 | 0.4 | 3,257 | 4.3 |
| 362 | 476 | 0.6 | 3,733 | 4.9 |
| 363 | 620 | 0.8 | 4,353 | 5.7 |
| 364 | 328 | 0.4 | 4,681 | 6.1 |
| 365 | 1,124 | 1.5 | 5,805 | 7.6 |
| 366 | 366 | 0.5 | 6,171 | 8.1 |
| 367 | 1,200 | 1.6 | 7,371 | 9.7 |
| 368 | 541 | 0.7 | 7,912 | 10.4 |
| 369 | 957 | 1.3 | 8,869 | 11.6 |
| 370 | 1,174 | 1.5 | 10,043 | 13.2 |
| 371 | 682 | 0.9 | 10,725 | 14.0 |
| 372 | 1,615 | 2.1 | 12,340 | 16.2 |
| 373 | 287 | 0.4 | 12,627 | 16.5 |
| 374 | 1,896 | 2.5 | 14,523 | 19.0 |
| 375 | 272 | 0.4 | 14,795 | 19.4 |
| 376 | 1,901 | 2.5 | 16,696 | 21.9 |
| 377 | 439 | 0.6 | 17,135 | 22.4 |
| 378 | 1,468 | 1.9 | 18,603 | 24.4 |
| 379 | 852 | 1.1 | 19,455 | 25.5 |
| 380 | 1,301 | 1.7 | 20,756 | 27.2 |
| 381 | 976 | 1.3 | 21,732 | 28.5 |
| 382 | 1,074 | 1.4 | 22,806 | 29.9 |
| 383 | 1,186 | 1.6 | 23,992 | 31.4 |
| 384 | 884 | 1.2 | 24,876 | 32.6 |
| 385 | 1,292 | 1.7 | 26,168 | 34.3 |
| 386 | 716 | 0.9 | 26,884 | 35.2 |
| 387 | 1,440 | 1.9 | 28,324 | 37.1 |
| 388 | 731 | 1.0 | 29,055 | 38.0 |

Table 6.B.2 *(continuation one)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 389 | 1,326 | 1.7 | 30,381 | 39.8 |
| 390 | 766 | 1.0 | 31,147 | 40.8 |
| 391 | 1,526 | 2.0 | 32,673 | 42.8 |
| 392 | 556 | 0.7 | 33,229 | 43.5 |
| 393 | 1,535 | 2.0 | 34,764 | 45.5 |
| 394 | 696 | 0.9 | 35,460 | 46.4 |
| 395 | 1,479 | 1.9 | 36,939 | 48.4 |
| 396 | 1,229 | 1.6 | 38,168 | 50.0 |
| 397 | 981 | 1.3 | 39,149 | 51.3 |
| 398 | 1,005 | 1.3 | 40,154 | 52.6 |
| 399 | 945 | 1.2 | 41,099 | 53.8 |
| 400 | 1,093 | 1.4 | 42,192 | 55.2 |
| 401 | 913 | 1.2 | 43,105 | 56.4 |
| 402 | 960 | 1.3 | 44,065 | 57.7 |
| 403 | 1,539 | 2.0 | 45,604 | 59.7 |
| 404 | 1,048 | 1.4 | 46,652 | 61.1 |
| 405 | 948 | 1.2 | 47,600 | 62.3 |
| 406 | 882 | 1.2 | 48,482 | 63.5 |
| 407 | 885 | 1.2 | 49,367 | 64.6 |
| 408 | 922 | 1.2 | 50,289 | 65.8 |
| 409 | 997 | 1.3 | 51,286 | 67.2 |
| 410 | 983 | 1.3 | 52,269 | 68.4 |
| 411 | 1,126 | 1.5 | 53,395 | 69.9 |
| 412 | 960 | 1.3 | 54,355 | 71.2 |
| 413 | 705 | 0.9 | 55,060 | 72.1 |
| 414 | 1,044 | 1.4 | 56,104 | 73.5 |
| 415 | 941 | 1.2 | 57,045 | 74.7 |
| 416 | 875 | 1.1 | 57,920 | 75.8 |
| 417 | 749 | 1.0 | 58,669 | 76.8 |
| 418 | 748 | 1.0 | 59,417 | 77.8 |
| 419 | 1,202 | 1.6 | 60,619 | 79.4 |
| 420 | 594 | 0.8 | 61,213 | 80.2 |
| 421 | 1,085 | 1.4 | 62,298 | 81.6 |
| 422 | 464 | 0.6 | 62,762 | 82.2 |
| 423 | 990 | 1.3 | 63,752 | 83.5 |
| 424 | 648 | 0.8 | 64,400 | 84.3 |
| 425 | 676 | 0.9 | 65,076 | 85.2 |
| 426 | 799 | 1.0 | 65,875 | 86.3 |

Table 6.B.2 *(continuation two)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 427 | 578 | 0.8 | 66,453 | 87.0 |
| 428 | 741 | 1.0 | 67,194 | 88.0 |
| 429 | 558 | 0.7 | 67,752 | 88.7 |
| 430 | 807 | 1.1 | 68,559 | 89.8 |
| 431 | 612 | 0.8 | 69,171 | 90.6 |
| 432 | 511 | 0.7 | 69,682 | 91.2 |
| 433 | 631 | 0.8 | 70,313 | 92.1 |
| 434 | 541 | 0.7 | 70,854 | 92.8 |
| 435 | 440 | 0.6 | 71,294 | 93.4 |
| 436 | 560 | 0.7 | 71,854 | 94.1 |
| 437 | 429 | 0.6 | 72,283 | 94.6 |
| 438 | 518 | 0.7 | 72,801 | 95.3 |
| 439 | 478 | 0.6 | 73,279 | 96.0 |
| 440 | 402 | 0.5 | 73,681 | 96.5 |
| 441 | 386 | 0.5 | 74,067 | 97.0 |
| 442 | 267 | 0.3 | 74,334 | 97.3 |
| 443 | 417 | 0.5 | 74,751 | 97.9 |
| 444 | 379 | 0.5 | 75,130 | 98.4 |
| 445 | 230 | 0.3 | 75,360 | 98.7 |
| 446 | 262 | 0.3 | 75,622 | 99.0 |
| 447 | 259 | 0.3 | 75,881 | 99.4 |
| 448 | 180 | 0.2 | 76,061 | 99.6 |
| 449 | 132 | 0.2 | 76,193 | 99.8 |
| 450 | 177 | 0.2 | 76,370 | 100.0 |

Table 6.B.3 Overall Scale Score Distribution for Grade Ten

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 550 | 7 | 0.3 | 7 | 0.3 |
| 551 | 0 | 0.0 | 7 | 0.3 |
| 552 | 0 | 0.0 | 7 | 0.3 |
| 553 | 0 | 0.0 | 7 | 0.3 |
| 554 | 1 | 0.0 | 8 | 0.4 |
| 555 | 0 | 0.0 | 8 | 0.4 |
| 556 | 0 | 0.0 | 8 | 0.4 |
| 557 | 2 | 0.1 | 10 | 0.4 |
| 558 | 3 | 0.1 | 13 | 0.6 |
| 559 | 5 | 0.2 | 18 | 0.8 |
| 560 | 0 | 0.0 | 18 | 0.8 |
| 561 | 7 | 0.3 | 25 | 1.1 |
| 562 | 0 | 0.0 | 25 | 1.1 |
| 563 | 5 | 0.2 | 30 | 1.3 |
| 564 | 13 | 0.6 | 43 | 1.9 |
| 565 | 12 | 0.5 | 55 | 2.4 |
| 566 | 7 | 0.3 | 62 | 2.7 |
| 567 | 13 | 0.6 | 75 | 3.3 |
| 568 | 28 | 1.2 | 103 | 4.5 |
| 569 | 16 | 0.7 | 119 | 5.2 |
| 570 | 30 | 1.3 | 149 | 6.5 |
| 571 | 13 | 0.6 | 162 | 7.1 |
| 572 | 22 | 1.0 | 184 | 8.1 |
| 573 | 16 | 0.7 | 200 | 8.8 |
| 574 | 27 | 1.2 | 227 | 10.0 |
| 575 | 16 | 0.7 | 243 | 10.7 |
| 576 | 24 | 1.1 | 267 | 11.7 |
| 577 | 36 | 1.6 | 303 | 13.3 |
| 578 | 7 | 0.3 | 310 | 13.6 |
| 579 | 51 | 2.2 | 361 | 15.9 |
| 580 | 13 | 0.6 | 374 | 16.4 |
| 581 | 50 | 2.2 | 424 | 18.6 |
| 582 | 10 | 0.4 | 434 | 19.1 |
| 583 | 29 | 1.3 | 463 | 20.3 |
| 584 | 34 | 1.5 | 497 | 21.8 |
| 585 | 25 | 1.1 | 522 | 22.9 |
| 586 | 42 | 1.8 | 564 | 24.8 |
| 587 | 22 | 1.0 | 586 | 25.7 |
| 588 | 53 | 2.3 | 639 | 28.1 |

Table 6.B.3 *(continuation one)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 589 | 24 | 1.1 | 663 | 29.1 |
| 590 | 33 | 1.4 | 696 | 30.6 |
| 591 | 20 | 0.9 | 716 | 31.4 |
| 592 | 44 | 1.9 | 760 | 33.4 |
| 593 | 26 | 1.1 | 786 | 34.5 |
| 594 | 43 | 1.9 | 829 | 36.4 |
| 595 | 19 | 0.8 | 848 | 37.2 |
| 596 | 35 | 1.5 | 883 | 38.8 |
| 597 | 34 | 1.5 | 917 | 40.3 |
| 598 | 34 | 1.5 | 951 | 41.8 |
| 599 | 32 | 1.4 | 983 | 43.2 |
| 600 | 34 | 1.5 | 1,017 | 44.7 |
| 601 | 33 | 1.4 | 1,050 | 46.1 |
| 602 | 30 | 1.3 | 1,080 | 47.4 |
| 603 | 23 | 1.0 | 1,103 | 48.4 |
| 604 | 43 | 1.9 | 1,146 | 50.3 |
| 605 | 22 | 1.0 | 1,168 | 51.3 |
| 606 | 44 | 1.9 | 1,212 | 53.2 |
| 607 | 12 | 0.5 | 1,224 | 53.8 |
| 608 | 54 | 2.4 | 1,278 | 56.1 |
| 609 | 17 | 0.7 | 1,295 | 56.9 |
| 610 | 58 | 2.5 | 1,353 | 59.4 |
| 611 | 13 | 0.6 | 1,366 | 60.0 |
| 612 | 43 | 1.9 | 1,409 | 61.9 |
| 613 | 34 | 1.5 | 1,443 | 63.4 |
| 614 | 44 | 1.9 | 1,487 | 65.3 |
| 615 | 22 | 1.0 | 1,509 | 66.3 |
| 616 | 38 | 1.7 | 1,547 | 67.9 |
| 617 | 15 | 0.7 | 1,562 | 68.6 |
| 618 | 35 | 1.5 | 1,597 | 70.1 |
| 619 | 24 | 1.1 | 1,621 | 71.2 |
| 620 | 34 | 1.5 | 1,655 | 72.7 |
| 621 | 21 | 0.9 | 1,676 | 73.6 |
| 622 | 43 | 1.9 | 1,719 | 75.5 |
| 623 | 14 | 0.6 | 1,733 | 76.1 |
| 624 | 37 | 1.6 | 1,770 | 77.7 |
| 625 | 18 | 0.8 | 1,788 | 78.5 |
| 626 | 23 | 1.0 | 1,811 | 79.5 |

Table 6.B.3 *(continuation two)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 627 | 20 | 0.9 | 1,831 | 80.4 |
| 628 | 28 | 1.2 | 1,859 | 81.6 |
| 629 | 25 | 1.1 | 1,884 | 82.7 |
| 630 | 31 | 1.4 | 1,915 | 84.1 |
| 631 | 19 | 0.8 | 1,934 | 84.9 |
| 632 | 21 | 0.9 | 1,955 | 85.9 |
| 633 | 20 | 0.9 | 1,975 | 86.7 |
| 634 | 22 | 1.0 | 1,997 | 87.7 |
| 635 | 17 | 0.7 | 2,014 | 88.4 |
| 636 | 15 | 0.7 | 2,029 | 89.1 |
| 637 | 34 | 1.5 | 2,063 | 90.6 |
| 638 | 28 | 1.2 | 2,091 | 91.8 |
| 639 | 8 | 0.4 | 2,099 | 92.2 |
| 640 | 25 | 1.1 | 2,124 | 93.3 |
| 641 | 17 | 0.7 | 2,141 | 94.0 |
| 642 | 9 | 0.4 | 2,150 | 94.4 |
| 643 | 29 | 1.3 | 2,179 | 95.7 |
| 644 | 29 | 1.3 | 2,208 | 97.0 |
| 645 | 5 | 0.2 | 2,213 | 97.2 |
| 646 | 18 | 0.8 | 2,231 | 98.0 |
| 647 | 17 | 0.7 | 2,248 | 98.7 |
| 648 | 13 | 0.6 | 2,261 | 99.3 |
| 649 | 4 | 0.2 | 2,265 | 99.5 |
| 650 | 12 | 0.5 | 2,277 | 100.0 |

Table 6.B.4 Overall Scale Score Distribution for Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 550 | 370 | 0.8 | 370 | 0.8 |
| 551 | 1 | 0.0 | 371 | 0.8 |
| 552 | 7 | 0.0 | 378 | 0.9 |
| 553 | 10 | 0.0 | 388 | 0.9 |
| 554 | 28 | 0.1 | 416 | 0.9 |
| 555 | 35 | 0.1 | 451 | 1.0 |
| 556 | 51 | 0.1 | 502 | 1.1 |
| 557 | 40 | 0.1 | 542 | 1.2 |
| 558 | 78 | 0.2 | 620 | 1.4 |
| 559 | 141 | 0.3 | 761 | 1.7 |
| 560 | 18 | 0.0 | 779 | 1.8 |
| 561 | 328 | 0.7 | 1,107 | 2.5 |
| 562 | 40 | 0.1 | 1,147 | 2.6 |
| 563 | 255 | 0.6 | 1,402 | 3.2 |
| 564 | 297 | 0.7 | 1,699 | 3.8 |
| 565 | 295 | 0.7 | 1,994 | 4.5 |
| 566 | 265 | 0.6 | 2,259 | 5.1 |
| 567 | 341 | 0.8 | 2,600 | 5.9 |
| 568 | 505 | 1.1 | 3,105 | 7.0 |
| 569 | 279 | 0.6 | 3,384 | 7.7 |
| 570 | 610 | 1.4 | 3,994 | 9.0 |
| 571 | 311 | 0.7 | 4,305 | 9.7 |
| 572 | 543 | 1.2 | 4,848 | 11.0 |
| 573 | 429 | 1.0 | 5,277 | 11.9 |
| 574 | 701 | 1.6 | 5,978 | 13.5 |
| 575 | 467 | 1.1 | 6,445 | 14.6 |
| 576 | 519 | 1.2 | 6,964 | 15.8 |
| 577 | 720 | 1.6 | 7,684 | 17.4 |
| 578 | 184 | 0.4 | 7,868 | 17.8 |
| 579 | 927 | 2.1 | 8,795 | 19.9 |
| 580 | 294 | 0.7 | 9,089 | 20.6 |
| 581 | 936 | 2.1 | 10,025 | 22.7 |
| 582 | 329 | 0.7 | 10,354 | 23.4 |
| 583 | 686 | 1.6 | 11,040 | 25.0 |
| 584 | 662 | 1.5 | 11,702 | 26.5 |
| 585 | 466 | 1.1 | 12,168 | 27.6 |
| 586 | 825 | 1.9 | 12,993 | 29.4 |
| 587 | 310 | 0.7 | 13,303 | 30.1 |
| 588 | 839 | 1.9 | 14,142 | 32.0 |

Table 6.B.4 *(continuation one)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 589 | 486 | 1.1 | 14,628 | 33.1 |
| 590 | 699 | 1.6 | 15,327 | 34.7 |
| 591 | 516 | 1.2 | 15,843 | 35.9 |
| 592 | 763 | 1.7 | 16,606 | 37.6 |
| 593 | 466 | 1.1 | 17,072 | 38.7 |
| 594 | 823 | 1.9 | 17,895 | 40.5 |
| 595 | 472 | 1.1 | 18,367 | 41.6 |
| 596 | 653 | 1.5 | 19,020 | 43.1 |
| 597 | 674 | 1.5 | 19,694 | 44.6 |
| 598 | 653 | 1.5 | 20,347 | 46.1 |
| 599 | 489 | 1.1 | 20,836 | 47.2 |
| 600 | 824 | 1.9 | 21,660 | 49.0 |
| 601 | 505 | 1.1 | 22,165 | 50.2 |
| 602 | 625 | 1.4 | 22,790 | 51.6 |
| 603 | 512 | 1.2 | 23,302 | 52.8 |
| 604 | 859 | 1.9 | 24,161 | 54.7 |
| 605 | 464 | 1.1 | 24,625 | 55.8 |
| 606 | 820 | 1.9 | 25,445 | 57.6 |
| 607 | 321 | 0.7 | 25,766 | 58.3 |
| 608 | 949 | 2.1 | 26,715 | 60.5 |
| 609 | 321 | 0.7 | 27,036 | 61.2 |
| 610 | 962 | 2.2 | 27,998 | 63.4 |
| 611 | 323 | 0.7 | 28,321 | 64.1 |
| 612 | 827 | 1.9 | 29,148 | 66.0 |
| 613 | 440 | 1.0 | 29,588 | 67.0 |
| 614 | 782 | 1.8 | 30,370 | 68.8 |
| 615 | 438 | 1.0 | 30,808 | 69.8 |
| 616 | 704 | 1.6 | 31,512 | 71.4 |
| 617 | 456 | 1.0 | 31,968 | 72.4 |
| 618 | 709 | 1.6 | 32,677 | 74.0 |
| 619 | 421 | 1.0 | 33,098 | 74.9 |
| 620 | 851 | 1.9 | 33,949 | 76.9 |
| 621 | 440 | 1.0 | 34,389 | 77.9 |
| 622 | 814 | 1.8 | 35,203 | 79.7 |
| 623 | 242 | 0.5 | 35,445 | 80.3 |
| 624 | 632 | 1.4 | 36,077 | 81.7 |
| 625 | 363 | 0.8 | 36,440 | 82.5 |
| 626 | 600 | 1.4 | 37,040 | 83.9 |

Table 6.B.4 *(continuation two)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 627 | 355 | 0.8 | 37,395 | 84.7 |
| 628 | 569 | 1.3 | 37,964 | 86.0 |
| 629 | 417 | 0.9 | 38,381 | 86.9 |
| 630 | 643 | 1.5 | 39,024 | 88.4 |
| 631 | 276 | 0.6 | 39,300 | 89.0 |
| 632 | 492 | 1.1 | 39,792 | 90.1 |
| 633 | 375 | 0.8 | 40,167 | 91.0 |
| 634 | 425 | 1.0 | 40,592 | 91.9 |
| 635 | 407 | 0.9 | 40,999 | 92.8 |
| 636 | 327 | 0.7 | 41,326 | 93.6 |
| 637 | 396 | 0.9 | 41,722 | 94.5 |
| 638 | 329 | 0.7 | 42,051 | 95.2 |
| 639 | 199 | 0.5 | 42,250 | 95.7 |
| 640 | 382 | 0.9 | 42,632 | 96.5 |
| 641 | 284 | 0.6 | 42,916 | 97.2 |
| 642 | 71 | 0.2 | 42,987 | 97.3 |
| 643 | 302 | 0.7 | 43,289 | 98.0 |
| 644 | 240 | 0.5 | 43,529 | 98.6 |
| 645 | 58 | 0.1 | 43,587 | 98.7 |
| 646 | 155 | 0.4 | 43,742 | 99.1 |
| 647 | 169 | 0.4 | 43,911 | 99.4 |
| 648 | 97 | 0.2 | 44,008 | 99.7 |
| 649 | 56 | 0.1 | 44,064 | 99.8 |
| 650 | 97 | 0.2 | 44,161 | 100.0 |

Table 6.B.5 Overall Scale Score Distribution for Grade Twelve

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 550 | 416 | 0.7 | 416 | 0.7 |
| 551 | 6 | 0.0 | 422 | 0.7 |
| 552 | 11 | 0.0 | 433 | 0.7 |
| 553 | 21 | 0.0 | 454 | 0.7 |
| 554 | 26 | 0.0 | 480 | 0.8 |
| 555 | 42 | 0.1 | 522 | 0.8 |
| 556 | 95 | 0.2 | 617 | 1.0 |
| 557 | 54 | 0.1 | 671 | 1.1 |
| 558 | 95 | 0.2 | 766 | 1.2 |
| 559 | 176 | 0.3 | 942 | 1.5 |
| 560 | 14 | 0.0 | 956 | 1.5 |
| 561 | 439 | 0.7 | 1,395 | 2.2 |
| 562 | 51 | 0.1 | 1,446 | 2.3 |
| 563 | 386 | 0.6 | 1,832 | 2.9 |
| 564 | 386 | 0.6 | 2,218 | 3.6 |
| 565 | 369 | 0.6 | 2,587 | 4.2 |
| 566 | 416 | 0.7 | 3,003 | 4.8 |
| 567 | 479 | 0.8 | 3,482 | 5.6 |
| 568 | 755 | 1.2 | 4,237 | 6.8 |
| 569 | 425 | 0.7 | 4,662 | 7.5 |
| 570 | 938 | 1.5 | 5,600 | 9.0 |
| 571 | 439 | 0.7 | 6,039 | 9.7 |
| 572 | 823 | 1.3 | 6,862 | 11.0 |
| 573 | 643 | 1.0 | 7,505 | 12.1 |
| 574 | 1,066 | 1.7 | 8,571 | 13.8 |
| 575 | 684 | 1.1 | 9,255 | 14.9 |
| 576 | 749 | 1.2 | 10,004 | 16.1 |
| 577 | 1,163 | 1.9 | 11,167 | 18.0 |
| 578 | 288 | 0.5 | 11,455 | 18.4 |
| 579 | 1,430 | 2.3 | 12,885 | 20.7 |
| 580 | 488 | 0.8 | 13,373 | 21.5 |
| 581 | 1,498 | 2.4 | 14,871 | 23.9 |
| 582 | 511 | 0.8 | 15,382 | 24.8 |
| 583 | 978 | 1.6 | 16,360 | 26.3 |
| 584 | 1,020 | 1.6 | 17,380 | 28.0 |
| 585 | 668 | 1.1 | 18,048 | 29.0 |
| 586 | 1,261 | 2.0 | 19,309 | 31.1 |
| 587 | 525 | 0.8 | 19,834 | 31.9 |
| 588 | 1,346 | 2.2 | 21,180 | 34.1 |

Table 6.B.5 *(continuation one)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 589 | 745 | 1.2 | 21,925 | 35.3 |
| 590 | 991 | 1.6 | 22,916 | 36.9 |
| 591 | 801 | 1.3 | 23,717 | 38.2 |
| 592 | 1,239 | 2.0 | 24,956 | 40.2 |
| 593 | 809 | 1.3 | 25,765 | 41.5 |
| 594 | 1,242 | 2.0 | 27,007 | 43.5 |
| 595 | 749 | 1.2 | 27,756 | 44.7 |
| 596 | 1,060 | 1.7 | 28,816 | 46.4 |
| 597 | 1,017 | 1.6 | 29,833 | 48.0 |
| 598 | 920 | 1.5 | 30,753 | 49.5 |
| 599 | 715 | 1.2 | 31,468 | 50.6 |
| 600 | 1,259 | 2.0 | 32,727 | 52.7 |
| 601 | 718 | 1.2 | 33,445 | 53.8 |
| 602 | 971 | 1.6 | 34,416 | 55.4 |
| 603 | 728 | 1.2 | 35,144 | 56.6 |
| 604 | 1,190 | 1.9 | 36,334 | 58.5 |
| 605 | 714 | 1.1 | 37,048 | 59.6 |
| 606 | 1,171 | 1.9 | 38,219 | 61.5 |
| 607 | 479 | 0.8 | 38,698 | 62.3 |
| 608 | 1,372 | 2.2 | 40,070 | 64.5 |
| 609 | 420 | 0.7 | 40,490 | 65.2 |
| 610 | 1,402 | 2.3 | 41,892 | 67.4 |
| 611 | 448 | 0.7 | 42,340 | 68.1 |
| 612 | 1,109 | 1.8 | 43,449 | 69.9 |
| 613 | 616 | 1.0 | 44,065 | 70.9 |
| 614 | 1,075 | 1.7 | 45,140 | 72.6 |
| 615 | 600 | 1.0 | 45,740 | 73.6 |
| 616 | 1,086 | 1.7 | 46,826 | 75.4 |
| 617 | 577 | 0.9 | 47,403 | 76.3 |
| 618 | 977 | 1.6 | 48,380 | 77.9 |
| 619 | 540 | 0.9 | 48,920 | 78.7 |
| 620 | 1,091 | 1.8 | 50,011 | 80.5 |
| 621 | 542 | 0.9 | 50,553 | 81.4 |
| 622 | 1,048 | 1.7 | 51,601 | 83.0 |
| 623 | 315 | 0.5 | 51,916 | 83.5 |
| 624 | 810 | 1.3 | 52,726 | 84.9 |
| 625 | 466 | 0.7 | 53,192 | 85.6 |
| 626 | 786 | 1.3 | 53,978 | 86.9 |

Table 6.B.5 *(continuation two)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 627 | 428 | 0.7 | 54,406 | 87.6 |
| 628 | 709 | 1.1 | 55,115 | 88.7 |
| 629 | 520 | 0.8 | 55,635 | 89.5 |
| 630 | 759 | 1.2 | 56,394 | 90.8 |
| 631 | 339 | 0.5 | 56,733 | 91.3 |
| 632 | 534 | 0.9 | 57,267 | 92.2 |
| 633 | 406 | 0.7 | 57,673 | 92.8 |
| 634 | 522 | 0.8 | 58,195 | 93.7 |
| 635 | 423 | 0.7 | 58,618 | 94.3 |
| 636 | 368 | 0.6 | 58,986 | 94.9 |
| 637 | 472 | 0.8 | 59,458 | 95.7 |
| 638 | 350 | 0.6 | 59,808 | 96.2 |
| 639 | 207 | 0.3 | 60,015 | 96.6 |
| 640 | 434 | 0.7 | 60,449 | 97.3 |
| 641 | 308 | 0.5 | 60,757 | 97.8 |
| 642 | 109 | 0.2 | 60,866 | 98.0 |
| 643 | 339 | 0.5 | 61,205 | 98.5 |
| 644 | 245 | 0.4 | 61,450 | 98.9 |
| 645 | 57 | 0.1 | 61,507 | 99.0 |
| 646 | 176 | 0.3 | 61,683 | 99.3 |
| 647 | 170 | 0.3 | 61,853 | 99.5 |
| 648 | 123 | 0.2 | 61,976 | 99.7 |
| 649 | 47 | 0.1 | 62,023 | 99.8 |
| 650 | 116 | 0.2 | 62,139 | 100.0 |

Table 6.B.6 Overall Scale Score Distribution for High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 550 | 793 | 0.7 | 793 | 0.7 |
| 551 | 7 | 0.0 | 800 | 0.7 |
| 552 | 18 | 0.0 | 818 | 0.8 |
| 553 | 31 | 0.0 | 849 | 0.8 |
| 554 | 55 | 0.1 | 904 | 0.8 |
| 555 | 77 | 0.1 | 981 | 0.9 |
| 556 | 146 | 0.1 | 1,127 | 1.0 |
| 557 | 96 | 0.1 | 1,223 | 1.1 |
| 558 | 176 | 0.2 | 1,399 | 1.3 |
| 559 | 322 | 0.3 | 1,721 | 1.6 |
| 560 | 32 | 0.0 | 1,753 | 1.6 |
| 561 | 774 | 0.7 | 2,527 | 2.3 |
| 562 | 91 | 0.1 | 2,618 | 2.4 |
| 563 | 646 | 0.6 | 3,264 | 3.0 |
| 564 | 696 | 0.6 | 3,960 | 3.6 |
| 565 | 676 | 0.6 | 4,636 | 4.3 |
| 566 | 688 | 0.6 | 5,324 | 4.9 |
| 567 | 833 | 0.8 | 6,157 | 5.7 |
| 568 | 1,288 | 1.2 | 7,445 | 6.9 |
| 569 | 720 | 0.7 | 8,165 | 7.5 |
| 570 | 1,578 | 1.5 | 9,743 | 9.0 |
| 571 | 763 | 0.7 | 10,506 | 9.7 |
| 572 | 1,388 | 1.3 | 11,894 | 11.0 |
| 573 | 1,088 | 1.0 | 12,982 | 12.0 |
| 574 | 1,794 | 1.7 | 14,776 | 13.6 |
| 575 | 1,167 | 1.1 | 15,943 | 14.7 |
| 576 | 1,292 | 1.2 | 17,235 | 15.9 |
| 577 | 1,919 | 1.8 | 19,154 | 17.6 |
| 578 | 479 | 0.4 | 19,633 | 18.1 |
| 579 | 2,408 | 2.2 | 22,041 | 20.3 |
| 580 | 795 | 0.7 | 22,836 | 21.0 |
| 581 | 2,484 | 2.3 | 25,320 | 23.3 |
| 582 | 850 | 0.8 | 26,170 | 24.1 |
| 583 | 1,693 | 1.6 | 27,863 | 25.7 |
| 584 | 1,716 | 1.6 | 29,579 | 27.2 |
| 585 | 1,159 | 1.1 | 30,738 | 28.3 |
| 586 | 2,128 | 2.0 | 32,866 | 30.3 |
| 587 | 857 | 0.8 | 33,723 | 31.1 |
| 588 | 2,238 | 2.1 | 35,961 | 33.1 |

Table 6.B.6 *(continuation one)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 589 | 1,255 | 1.2 | 37,216 | 34.3 |
| 590 | 1,723 | 1.6 | 38,939 | 35.9 |
| 591 | 1,337 | 1.2 | 40,276 | 37.1 |
| 592 | 2,046 | 1.9 | 42,322 | 39.0 |
| 593 | 1,301 | 1.2 | 43,623 | 40.2 |
| 594 | 2,108 | 1.9 | 45,731 | 42.1 |
| 595 | 1,240 | 1.1 | 46,971 | 43.3 |
| 596 | 1,748 | 1.6 | 48,719 | 44.9 |
| 597 | 1,725 | 1.6 | 50,444 | 46.5 |
| 598 | 1,607 | 1.5 | 52,051 | 47.9 |
| 599 | 1,236 | 1.1 | 53,287 | 49.1 |
| 600 | 2,117 | 1.9 | 55,404 | 51.0 |
| 601 | 1,256 | 1.2 | 56,660 | 52.2 |
| 602 | 1,626 | 1.5 | 58,286 | 53.7 |
| 603 | 1,263 | 1.2 | 59,549 | 54.8 |
| 604 | 2,092 | 1.9 | 61,641 | 56.8 |
| 605 | 1,200 | 1.1 | 62,841 | 57.9 |
| 606 | 2,035 | 1.9 | 64,876 | 59.8 |
| 607 | 812 | 0.7 | 65,688 | 60.5 |
| 608 | 2,375 | 2.2 | 68,063 | 62.7 |
| 609 | 758 | 0.7 | 68,821 | 63.4 |
| 610 | 2,422 | 2.2 | 71,243 | 65.6 |
| 611 | 784 | 0.7 | 72,027 | 66.3 |
| 612 | 1,979 | 1.8 | 74,006 | 68.2 |
| 613 | 1,090 | 1.0 | 75,096 | 69.2 |
| 614 | 1,901 | 1.8 | 76,997 | 70.9 |
| 615 | 1,060 | 1.0 | 78,057 | 71.9 |
| 616 | 1,828 | 1.7 | 79,885 | 73.6 |
| 617 | 1,048 | 1.0 | 80,933 | 74.5 |
| 618 | 1,721 | 1.6 | 82,654 | 76.1 |
| 619 | 985 | 0.9 | 83,639 | 77.0 |
| 620 | 1,976 | 1.8 | 85,615 | 78.9 |
| 621 | 1,003 | 0.9 | 86,618 | 79.8 |
| 622 | 1,905 | 1.8 | 88,523 | 81.5 |
| 623 | 571 | 0.5 | 89,094 | 82.1 |
| 624 | 1,479 | 1.4 | 90,573 | 83.4 |
| 625 | 847 | 0.8 | 91,420 | 84.2 |
| 626 | 1,409 | 1.3 | 92,829 | 85.5 |
| 627 | 803 | 0.7 | 93,632 | 86.2 |

Table 6.B.6 *(continuation two)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scale Score | N | Percent | Cumulative Frequency | Cumulative Percent |
| 628 | 1,306 | 1.2 | 94,938 | 87.4 |
| 629 | 962 | 0.9 | 95,900 | 88.3 |
| 630 | 1,433 | 1.3 | 97,333 | 89.6 |
| 631 | 634 | 0.6 | 97,967 | 90.2 |
| 632 | 1,047 | 1.0 | 99,014 | 91.2 |
| 633 | 801 | 0.7 | 99,815 | 91.9 |
| 634 | 969 | 0.9 | 100,784 | 92.8 |
| 635 | 847 | 0.8 | 101,631 | 93.6 |
| 636 | 710 | 0.7 | 102,341 | 94.3 |
| 637 | 902 | 0.8 | 103,243 | 95.1 |
| 638 | 707 | 0.7 | 103,950 | 95.7 |
| 639 | 414 | 0.4 | 104,364 | 96.1 |
| 640 | 841 | 0.8 | 105,205 | 96.9 |
| 641 | 609 | 0.6 | 105,814 | 97.5 |
| 642 | 189 | 0.2 | 106,003 | 97.6 |
| 643 | 670 | 0.6 | 106,673 | 98.2 |
| 644 | 514 | 0.5 | 107,187 | 98.7 |
| 645 | 120 | 0.1 | 107,307 | 98.8 |
| 646 | 349 | 0.3 | 107,656 | 99.2 |
| 647 | 356 | 0.3 | 108,012 | 99.5 |
| 648 | 233 | 0.2 | 108,245 | 99.7 |
| 649 | 107 | 0.1 | 108,352 | 99.8 |
| 650 | 225 | 0.2 | 108,577 | 100.0 |

### Appendix 6.C: Demographic Summary of Overall Achievement Levels

**Note:** Student results are suppressed and indicted as “N/A” in table 6.C.1 through table 6.C.6 where fewer than 11 students are reported in a category.

Table 6.C.1 Percent of Students in Each Achievement Level for Total Scores by Demographic Variables for Grade Five

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | 62,238 | 199 | 22.3 | 21.9 | 48.3 | 19.8 | 9.9 | 29.7 |
| Male | 31,577 | 199 | 22.9 | 23.7 | 46.1 | 19.9 | 10.4 | 30.2 |
| Female | 30,659 | 200 | 21.7 | 20.1 | 50.7 | 19.8 | 9.5 | 29.2 |
| Nonbinary | 2 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| EL | 11,335 | 182 | 15.2 | 48.1 | 48.0 | 3.4 | 0.5 | 3.9 |
| English only | 38,862 | 202 | 22.0 | 18.0 | 47.9 | 22.8 | 11.2 | 34.1 |
| Reclassified fluent English proficient (RFEP) | 9,628 | 204 | 19.8 | 11.1 | 54.6 | 23.9 | 10.5 | 34.4 |
| IFEP | 2,372 | 218 | 20.5 | 5.2 | 31.2 | 31.5 | 32.0 | 63.5 |
| Adult English learner (ADEL) | 0 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| To be determined | 12 | 193 | 29.9 | 33.3 | 33.3 | 25.0 | 8.3 | 33.3 |
| English proficiency unknown | 29 | 200 | 23.3 | 24.1 | 34.5 | 34.5 | 6.9 | 41.4 |
| Economically disadvantaged | 34,089 | 191 | 19.5 | 31.5 | 53.3 | 12.1 | 3.1 | 15.2 |
| Not economically disadvantaged | 28,149 | 209 | 21.4 | 10.4 | 42.3 | 29.2 | 18.2 | 47.4 |

Table 6.C.1 *(continuation one)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| American Indian or Alaska Native (All) | 503 | 193 | 19.7 | 29.6 | 51.9 | 15.1 | 3.4 | 18.5 |
| Asian (All) | 7,115 | 216 | 21.4 | 6.9 | 31.5 | 31.3 | 30.3 | 61.6 |
| Native Hawaiian or Other Pacific Islander (All) | 237 | 194 | 20.8 | 23.6 | 57.4 | 13.5 | 5.5 | 19.0 |
| Filipino (All) | 1,128 | 207 | 19.9 | 9.9 | 48.3 | 29.4 | 12.3 | 41.8 |
| Hispanic or Latino (All) | 30,227 | 191 | 19.4 | 31.0 | 53.8 | 12.1 | 3.1 | 15.1 |
| Black or African American (All) | 2,256 | 189 | 19.9 | 37.3 | 48.2 | 11.3 | 3.1 | 14.5 |
| White (All) | 17,170 | 206 | 21.1 | 12.3 | 46.2 | 28.0 | 13.5 | 41.5 |
| Two or more races (All) | 3,602 | 206 | 22.2 | 14.0 | 44.2 | 26.4 | 15.4 | 41.8 |
| Special education services | 7,041 | 184 | 20.2 | 50.8 | 38.2 | 7.5 | 3.5 | 11.0 |
| No special education services | 55,197 | 201 | 21.8 | 18.2 | 49.6 | 21.4 | 10.8 | 32.1 |
| Migrant education | 626 | 185 | 17.5 | 43.3 | 49.2 | 6.1 | 1.4 | 7.5 |
| Not migrant education | 61,612 | 199 | 22.3 | 21.7 | 48.3 | 19.9 | 10.0 | 30.0 |
| Military | 688 | 202 | 20.5 | 15.1 | 52.6 | 24.6 | 7.7 | 32.3 |
| Not military | 61,550 | 199 | 22.3 | 22.0 | 48.3 | 19.8 | 10.0 | 29.7 |
| Homeless | 1,463 | 187 | 18.8 | 39.2 | 49.5 | 9.2 | 2.1 | 11.3 |
| Not homeless | 60,775 | 200 | 22.3 | 21.5 | 48.3 | 20.1 | 10.1 | 30.2 |

Table 6.C.1 *(continuation two)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 147 | 199 | 19.7 | 17.7 | 56.5 | 19.0 | 6.8 | 25.9 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 356 | 190 | 19.2 | 34.6 | 50.0 | 13.5 | 2.0 | 15.4 |
| Asian (Primary ethnicity—not economically disadvantaged) | 5,451 | 221 | 18.7 | 3.2 | 25.4 | 34.6 | 36.8 | 71.4 |
| Asian (Primary ethnicity—economically disadvantaged) | 1,664 | 200 | 21.5 | 18.8 | 51.5 | 20.4 | 9.3 | 29.7 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 100 | 203 | 19.6 | 9.0 | 60.0 | 23.0 | 8.0 | 31.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 137 | 188 | 19.4 | 34.3 | 55.5 | 6.6 | 3.6 | 10.2 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 785 | 209 | 19.7 | 7.8 | 45.1 | 32.9 | 14.3 | 47.1 |
| Filipino (Primary ethnicity—economically disadvantaged) | 343 | 201 | 19.4 | 14.9 | 55.7 | 21.6 | 7.9 | 29.4 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 7,330 | 199 | 20.6 | 19.4 | 53.8 | 20.1 | 6.7 | 26.8 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 22,897 | 189 | 18.3 | 34.8 | 53.9 | 9.5 | 1.9 | 11.4 |

Table 6.C.1 *(continuation three)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 682 | 197 | 21.0 | 24.5 | 50.7 | 19.1 | 5.7 | 24.8 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,574 | 186 | 18.4 | 42.9 | 47.1 | 8.0 | 2.0 | 10.0 |
| White (Primary ethnicity—not economically disadvantaged) | 11,348 | 211 | 20.0 | 7.6 | 42.4 | 32.6 | 17.4 | 50.0 |
| White (Primary ethnicity—economically disadvantaged) | 5,822 | 198 | 20.5 | 21.5 | 53.7 | 19.1 | 5.7 | 24.8 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,306 | 212 | 20.9 | 8.3 | 39.7 | 31.1 | 20.9 | 52.0 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,296 | 196 | 20.9 | 24.3 | 52.1 | 18.0 | 5.6 | 23.6 |

Table 6.C.2 Percent of Students in Each Achievement Level for Total Scores by Demographic Variables for Grade Eight

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | 75,000 | 399 | 22.8 | 21.0 | 52.0 | 18.1 | 8.9 | 27.0 |
| Male | 38,344 | 398 | 23.7 | 23.8 | 48.3 | 18.1 | 9.9 | 27.9 |
| Female | 36,633 | 399 | 21.7 | 18.1 | 55.8 | 18.2 | 7.9 | 26.1 |
| Nonbinary | 23 | 404 | 25.8 | 21.7 | 47.8 | 8.7 | 21.7 | 30.4 |
| EL | 10,521 | 379 | 13.6 | 49.8 | 48.6 | 1.4 | 0.2 | 1.6 |
| English only | 42,628 | 402 | 22.7 | 17.5 | 51.1 | 21.1 | 10.4 | 31.5 |
| RFEP | 18,470 | 400 | 20.7 | 15.3 | 59.0 | 18.8 | 6.9 | 25.8 |
| IFEP | 3,344 | 416 | 22.7 | 7.4 | 35.1 | 29.0 | 28.4 | 57.4 |
| ADEL | 0 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| To be determined | 6 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 31 | 395 | 19.4 | 19.4 | 64.5 | 12.9 | 3.2 | 16.1 |
| Economically disadvantaged | 41,609 | 391 | 19.8 | 29.0 | 56.9 | 11.5 | 2.6 | 14.1 |
| Not economically disadvantaged | 33,391 | 408 | 22.6 | 11.1 | 45.8 | 26.3 | 16.8 | 43.1 |

Table 6.C.2 *(continuation one)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| American Indian or Alaska Native (All) | 553 | 391 | 20.0 | 26.9 | 58.6 | 11.6 | 2.9 | 14.5 |
| Asian (All) | 7,826 | 418 | 21.8 | 6.5 | 31.4 | 31.7 | 30.4 | 62.1 |
| Native Hawaiian or Other Pacific Islander (All) | 268 | 393 | 19.4 | 24.3 | 59.3 | 13.8 | 2.6 | 16.4 |
| Filipino (All) | 1,807 | 407 | 20.4 | 9.3 | 50.9 | 28.2 | 11.6 | 39.8 |
| Hispanic or Latino (All) | 38,340 | 391 | 19.7 | 28.5 | 57.5 | 11.4 | 2.6 | 14.0 |
| Black or African American (All) | 2,717 | 389 | 19.3 | 33.2 | 55.2 | 9.7 | 1.9 | 11.6 |
| White (All) | 19,707 | 406 | 22.1 | 12.9 | 49.6 | 25.0 | 12.5 | 37.5 |
| Two or more races (All) | 3,782 | 406 | 22.9 | 13.2 | 47.3 | 25.0 | 14.6 | 39.5 |
| Special education services | 8,188 | 381 | 17.5 | 51.4 | 42.6 | 4.5 | 1.5 | 6.0 |
| No special education services | 66,812 | 401 | 22.4 | 17.3 | 53.1 | 19.8 | 9.8 | 29.6 |
| Migrant education | 668 | 386 | 18.1 | 35.2 | 56.7 | 6.7 | 1.3 | 8.1 |
| Not migrant education | 74,332 | 399 | 22.8 | 20.9 | 51.9 | 18.2 | 9.0 | 27.2 |
| Military | 1,049 | 403 | 21.5 | 14.2 | 52.6 | 23.1 | 10.1 | 33.2 |
| Nonmilitary | 73,951 | 399 | 22.8 | 21.1 | 51.9 | 18.0 | 8.9 | 26.9 |
| Homeless | 1,773 | 387 | 18.4 | 34.5 | 56.2 | 8.0 | 1.2 | 9.2 |
| Not homeless | 73,227 | 399 | 22.8 | 20.7 | 51.9 | 18.3 | 9.1 | 27.5 |

Table 6.C.2 *(continuation two)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 182 | 397 | 20.8 | 17.0 | 60.4 | 17.0 | 5.5 | 22.5 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 371 | 388 | 18.9 | 31.8 | 57.7 | 8.9 | 1.6 | 10.5 |
| Asian (Primary ethnicity—not economically disadvantaged) | 5,816 | 423 | 19.0 | 3.1 | 25.3 | 34.6 | 36.9 | 71.5 |
| Asian (Primary ethnicity—economically disadvantaged) | 2,010 | 403 | 22.9 | 16.1 | 49.2 | 23.1 | 11.6 | 34.7 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 119 | 399 | 20.4 | 18.5 | 58.0 | 19.3 | 4.2 | 23.5 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 149 | 389 | 17.4 | 28.9 | 60.4 | 9.4 | 1.3 | 10.7 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 1,233 | 410 | 20.2 | 7.9 | 47.2 | 30.5 | 14.4 | 44.8 |
| Filipino (Primary ethnicity—economically disadvantaged) | 574 | 402 | 19.9 | 12.2 | 58.7 | 23.3 | 5.7 | 29.1 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 9,276 | 398 | 21.1 | 19.2 | 56.7 | 18.0 | 6.1 | 24.1 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 29,064 | 389 | 18.7 | 31.5 | 57.7 | 9.3 | 1.5 | 10.8 |

Table 6.C.2 *(continuation three)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 868 | 396 | 20.1 | 21.0 | 58.8 | 16.7 | 3.6 | 20.3 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,849 | 385 | 18.0 | 39.0 | 53.5 | 6.4 | 1.1 | 7.5 |
| White (Primary ethnicity—not economically disadvantaged) | 13,428 | 410 | 21.5 | 9.1 | 46.3 | 28.4 | 16.1 | 44.6 |
| White (Primary ethnicity—economically disadvantaged) | 6,279 | 397 | 20.8 | 21.1 | 56.6 | 17.5 | 4.8 | 22.3 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,469 | 412 | 21.8 | 7.8 | 43.2 | 29.2 | 19.8 | 49.0 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,313 | 395 | 21.1 | 23.3 | 55.1 | 17.0 | 4.6 | 21.6 |

Table 6.C.3 Percent of Students in Each Achievement Level for Total Scores by Demographic Variables for Grade Ten

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | 2,270 | 605 | 22.5 | 10.4 | 54.8 | 23.2 | 11.6 | 34.8 |
| Male | 1,135 | 605 | 23.4 | 11.0 | 54.4 | 21.9 | 12.8 | 34.6 |
| Female | 1,134 | 605 | 21.7 | 9.8 | 55.3 | 24.5 | 10.4 | 34.9 |
| Nonbinary | 1 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| EL | 156 | 581 | 11.4 | 33.3 | 66.0 | 0.6 | 0.0 | 0.6 |
| English only | 1,448 | 607 | 21.4 | 8.3 | 53.8 | 27.2 | 10.7 | 37.9 |
| RFEP | 519 | 602 | 21.5 | 11.0 | 62.4 | 17.1 | 9.4 | 26.6 |
| IFEP | 147 | 622 | 23.5 | 4.8 | 25.9 | 29.3 | 40.1 | 69.4 |
| ADEL | 0 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| To be determined | 0 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 0 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Economically disadvantaged | 992 | 595 | 18.4 | 15.5 | 69.6 | 12.7 | 2.2 | 14.9 |
| Not economically disadvantaged | 1,278 | 613 | 22.4 | 6.4 | 43.3 | 31.4 | 18.9 | 50.2 |

Table 6.C.3 *(continuation one)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| American Indian or Alaska Native (All) | 22 | 593 | 15.9 | 13.6 | 68.2 | 18.2 | 0.0 | 18.2 |
| Asian (All) | 323 | 626 | 22.1 | 4.3 | 19.8 | 29.4 | 46.4 | 75.9 |
| Native Hawaiian or Other Pacific Islander (All) | 7 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Filipino (All) | 19 | 611 | 19.6 | 5.3 | 47.4 | 36.8 | 10.5 | 47.4 |
| Hispanic or Latino (All) | 890 | 594 | 17.9 | 15.7 | 69.6 | 13.1 | 1.6 | 14.7 |
| Black or African American (All) | 53 | 599 | 20.6 | 17.0 | 60.4 | 18.9 | 3.8 | 22.6 |
| White (All) | 831 | 608 | 20.2 | 6.9 | 54.3 | 30.7 | 8.2 | 38.9 |
| Two or more races (All) | 125 | 612 | 23.8 | 8.8 | 40.8 | 29.6 | 20.8 | 50.4 |
| Special education services | 226 | 587 | 18.9 | 33.2 | 57.1 | 8.4 | 1.3 | 9.7 |
| No special education services | 2,044 | 607 | 22.0 | 7.9 | 54.5 | 24.9 | 12.7 | 37.6 |
| Migrant education | 13 | 582 | 17.7 | 38.5 | 53.8 | 7.7 | 0.0 | 7.7 |
| Not migrant education | 2,257 | 605 | 22.5 | 10.2 | 54.8 | 23.3 | 11.7 | 35.0 |
| Military | 14 | 604 | 18.2 | 0.0 | 64.3 | 35.7 | 0.0 | 35.7 |
| Not military | 2,256 | 605 | 22.6 | 10.5 | 54.7 | 23.1 | 11.7 | 34.8 |
| Homeless | 28 | 586 | 13.6 | 21.4 | 78.6 | 0.0 | 0.0 | 0.0 |
| Not homeless | 2,242 | 605 | 22.5 | 10.3 | 54.5 | 23.5 | 11.7 | 35.2 |

Table 6.C.3 *(continuation two)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 9 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 13 | 588 | 14.4 | 15.4 | 76.9 | 7.7 | 0.0 | 7.7 |
| Asian (Primary ethnicity—not economically disadvantaged) | 292 | 628 | 20.2 | 3.1 | 17.1 | 30.5 | 49.3 | 79.8 |
| Asian (Primary ethnicity—economically disadvantaged) | 31 | 605 | 27.6 | 16.1 | 45.2 | 19.4 | 19.4 | 38.7 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 6 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Filipino (Primary ethnicity—not economically disadvantaged) | 16 | 615 | 16.6 | 0.0 | 50.0 | 37.5 | 12.5 | 50.0 |
| Filipino (Primary ethnicity—economically disadvantaged) | 3 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 276 | 600 | 19.2 | 9.8 | 65.2 | 21.4 | 3.6 | 25.0 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 614 | 592 | 16.7 | 18.4 | 71.5 | 9.4 | 0.7 | 10.1 |

Table 6.C.3 *(continuation three)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 26 | 605 | 22.3 | 19.2 | 42.3 | 30.8 | 7.7 | 38.5 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 27 | 593 | 17.2 | 14.8 | 77.8 | 7.4 | 0.0 | 7.4 |
| White (Primary ethnicity—not economically disadvantaged) | 567 | 611 | 20.1 | 5.6 | 48.1 | 35.8 | 10.4 | 46.2 |
| White (Primary ethnicity—economically disadvantaged) | 264 | 601 | 18.6 | 9.5 | 67.4 | 19.7 | 3.4 | 23.1 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 86 | 617 | 23.6 | 8.1 | 29.1 | 36.0 | 26.7 | 62.8 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 39 | 601 | 20.4 | 10.3 | 66.7 | 15.4 | 7.7 | 23.1 |

Table 6.C.4 Percent of Students in Each Achievement Level for Total Scores by Demographic Variables for Grade Eleven

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | 43,791 | 602 | 22.2 | 13.9 | 54.6 | 24.3 | 7.2 | 31.5 |
| Male | 21,859 | 601 | 23.1 | 16.0 | 52.2 | 23.8 | 8.0 | 31.8 |
| Female | 21,913 | 603 | 21.2 | 11.8 | 57.1 | 24.7 | 6.4 | 31.1 |
| Nonbinary | 19 | 607 | 24.5 | 10.5 | 42.1 | 31.6 | 15.8 | 47.4 |
| EL | 3,402 | 581 | 13.7 | 39.9 | 58.3 | 1.7 | 0.0 | 1.8 |
| English only | 24,168 | 604 | 22.3 | 12.2 | 51.8 | 27.6 | 8.4 | 36.0 |
| RFEP | 14,341 | 602 | 20.5 | 11.5 | 60.4 | 22.8 | 5.3 | 28.1 |
| IFEP | 1,871 | 614 | 22.4 | 6.6 | 39.9 | 33.4 | 20.1 | 53.5 |
| ADEL | 0 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| To be determined | 1 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 8 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Economically disadvantaged | 23,793 | 597 | 20.4 | 17.6 | 61.2 | 18.4 | 2.8 | 21.3 |
| Not economically disadvantaged | 19,998 | 608 | 22.5 | 9.5 | 46.9 | 31.2 | 12.4 | 43.7 |

Table 6.C.4 *(continuation one)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| American Indian or Alaska Native (All) | 304 | 596 | 20.4 | 17.4 | 60.9 | 18.1 | 3.6 | 21.7 |
| Asian (All) | 4,494 | 616 | 22.4 | 5.9 | 34.9 | 36.0 | 23.3 | 59.3 |
| Native Hawaiian or Other Pacific Islander (All) | 174 | 598 | 21.2 | 15.5 | 62.1 | 18.4 | 4.0 | 22.4 |
| Filipino (All) | 1,319 | 612 | 20.6 | 5.8 | 44.5 | 37.5 | 12.1 | 49.7 |
| Hispanic or Latino (All) | 21,606 | 596 | 19.9 | 17.6 | 62.7 | 17.3 | 2.4 | 19.6 |
| Black or African American (All) | 1,402 | 593 | 21.0 | 23.5 | 57.4 | 16.4 | 2.6 | 19.0 |
| White (All) | 12,710 | 606 | 22.1 | 10.4 | 49.3 | 30.9 | 9.5 | 40.3 |
| Two or more races (All) | 1,782 | 607 | 22.6 | 11.1 | 47.9 | 30.6 | 10.5 | 41.1 |
| Special education services | 3,724 | 583 | 17.1 | 39.3 | 54.5 | 5.4 | 0.8 | 6.3 |
| No special education services | 40,067 | 604 | 21.8 | 11.5 | 54.7 | 26.0 | 7.8 | 33.8 |
| Migrant education | 266 | 591 | 18.1 | 21.4 | 66.2 | 12.4 | 0.0 | 12.4 |
| Not migrant education | 43,525 | 602 | 22.2 | 13.8 | 54.6 | 24.3 | 7.3 | 31.6 |
| Military | 654 | 605 | 23.4 | 11.9 | 51.7 | 24.6 | 11.8 | 36.4 |
| Not military | 43,137 | 602 | 22.2 | 13.9 | 54.7 | 24.3 | 7.2 | 31.4 |
| Homeless | 1,375 | 592 | 19.5 | 22.5 | 62.0 | 14.3 | 1.2 | 15.4 |
| Not homeless | 42,416 | 602 | 22.2 | 13.6 | 54.4 | 24.6 | 7.4 | 32.0 |

Table 6.C.4 *(continuation two)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 139 | 599 | 19.9 | 13.7 | 61.9 | 19.4 | 5.0 | 24.5 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 165 | 594 | 20.6 | 20.6 | 60.0 | 17.0 | 2.4 | 19.4 |
| Asian (Primary ethnicity—not economically disadvantaged) | 2,749 | 622 | 21.1 | 3.9 | 26.7 | 38.5 | 30.9 | 69.4 |
| Asian (Primary ethnicity—economically disadvantaged) | 1,745 | 608 | 22.0 | 9.0 | 47.7 | 32.0 | 11.3 | 43.3 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 81 | 605 | 21.7 | 9.9 | 59.3 | 23.5 | 7.4 | 30.9 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 93 | 593 | 19.4 | 20.4 | 64.5 | 14.0 | 1.1 | 15.1 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 865 | 615 | 19.8 | 4.3 | 40.5 | 40.9 | 14.3 | 55.3 |
| Filipino (Primary ethnicity—economically disadvantaged) | 454 | 607 | 21.2 | 8.8 | 52.2 | 31.1 | 7.9 | 39.0 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 5,515 | 600 | 21.0 | 14.3 | 58.7 | 22.4 | 4.6 | 27.0 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 16,091 | 595 | 19.3 | 18.8 | 64.1 | 15.5 | 1.6 | 17.1 |

Table 6.C.4 *(continuation three)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 520 | 598 | 22.4 | 19.2 | 53.3 | 22.1 | 5.4 | 27.5 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 882 | 591 | 19.7 | 26.1 | 59.9 | 13.0 | 1.0 | 14.1 |
| White (Primary ethnicity—not economically disadvantaged) | 8,982 | 609 | 21.8 | 8.3 | 46.0 | 33.9 | 11.8 | 45.6 |
| White (Primary ethnicity—economically disadvantaged) | 3,728 | 599 | 21.2 | 15.2 | 57.2 | 23.7 | 3.9 | 27.6 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 1,147 | 611 | 22.2 | 7.5 | 43.9 | 34.8 | 13.9 | 48.6 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 635 | 600 | 21.6 | 17.5 | 55.1 | 23.0 | 4.4 | 27.4 |

Table 6.C.5 Percent of Students in Each Achievement Level for Total Scores by Demographic Variables for Grade Twelve

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | 61,723 | 600 | 21.4 | 14.3 | 58.1 | 21.8 | 5.7 | 27.5 |
| Male | 30,416 | 600 | 22.4 | 16.3 | 55.1 | 22.1 | 6.6 | 28.7 |
| Female | 31,285 | 600 | 20.4 | 12.4 | 61.1 | 21.6 | 4.8 | 26.4 |
| Nonbinary | 22 | 604 | 25.7 | 13.6 | 45.5 | 31.8 | 9.1 | 40.9 |
| EL | 5,086 | 581 | 13.3 | 37.9 | 60.6 | 1.4 | 0.0 | 1.4 |
| English only | 30,988 | 603 | 21.9 | 12.7 | 54.3 | 25.6 | 7.4 | 33.0 |
| RFEP | 22,485 | 599 | 19.6 | 12.1 | 64.1 | 20.1 | 3.6 | 23.7 |
| IFEP | 3,086 | 609 | 22.1 | 7.7 | 48.2 | 30.3 | 13.8 | 44.1 |
| ADEL | 69 | 586 | 11.0 | 17.4 | 82.6 | 0.0 | 0.0 | 0.0 |
| To be determined | 2 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 7 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Economically disadvantaged | 37,023 | 596 | 19.6 | 17.3 | 63.9 | 16.5 | 2.3 | 18.8 |
| Not economically disadvantaged | 24,700 | 607 | 22.2 | 9.8 | 49.6 | 29.8 | 10.9 | 40.6 |

Table 6.C.5 *(continuation one)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| American Indian or Alaska Native (All) | 321 | 597 | 20.6 | 17.1 | 59.8 | 19.6 | 3.4 | 23.1 |
| Asian (All) | 5,296 | 614 | 22.1 | 5.8 | 39.3 | 35.0 | 19.9 | 54.9 |
| Native Hawaiian or Other Pacific Islander (All) | 241 | 597 | 19.2 | 16.6 | 64.3 | 17.4 | 1.7 | 19.1 |
| Filipino (All) | 1,916 | 608 | 20.0 | 6.1 | 52.7 | 33.0 | 8.2 | 41.2 |
| Hispanic or Latino (All) | 35,846 | 595 | 19.3 | 17.1 | 64.5 | 16.4 | 2.0 | 18.4 |
| Black or African American (All) | 2,419 | 591 | 19.7 | 25.3 | 60.4 | 12.4 | 1.8 | 14.2 |
| White (All) | 13,436 | 607 | 21.9 | 10.0 | 50.1 | 30.0 | 9.9 | 39.9 |
| Two or more races (All) | 2,248 | 606 | 22.3 | 11.3 | 49.3 | 30.3 | 9.1 | 39.5 |
| Special education services | 5,660 | 585 | 17.3 | 34.5 | 58.1 | 6.6 | 0.7 | 7.4 |
| No special education services | 56,063 | 602 | 21.2 | 12.3 | 58.1 | 23.4 | 6.2 | 29.6 |
| Migrant education | 474 | 590 | 17.9 | 21.5 | 67.5 | 10.3 | 0.6 | 11.0 |
| Not migrant education | 61,249 | 600 | 21.4 | 14.3 | 58.1 | 21.9 | 5.7 | 27.7 |
| Military | 960 | 601 | 21.9 | 14.1 | 54.3 | 25.9 | 5.7 | 31.7 |
| Not military | 60,763 | 600 | 21.4 | 14.3 | 58.2 | 21.8 | 5.7 | 27.5 |
| Homeless | 2,148 | 593 | 19.1 | 20.9 | 64.3 | 13.5 | 1.4 | 14.9 |
| Not homeless | 59,575 | 600 | 21.5 | 14.1 | 57.9 | 22.1 | 5.9 | 28.0 |

Table 6.C.5 *(continuation two)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 116 | 600 | 20.7 | 13.8 | 57.8 | 24.1 | 4.3 | 28.4 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 205 | 595 | 20.4 | 19.0 | 61.0 | 17.1 | 2.9 | 20.0 |
| Asian (Primary ethnicity—not economically disadvantaged) | 3,199 | 619 | 21.2 | 4.1 | 31.6 | 38.3 | 26.0 | 64.3 |
| Asian (Primary ethnicity—economically disadvantaged) | 2,097 | 607 | 21.7 | 8.3 | 51.1 | 29.9 | 10.6 | 40.6 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 109 | 599 | 18.9 | 11.9 | 70.6 | 13.8 | 3.7 | 17.4 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 132 | 596 | 19.3 | 20.5 | 59.1 | 20.5 | 0.0 | 20.5 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 1,197 | 610 | 19.8 | 4.9 | 50.5 | 34.5 | 10.1 | 44.6 |
| Filipino (Primary ethnicity—economically disadvantaged) | 719 | 605 | 19.9 | 8.1 | 56.5 | 30.5 | 5.0 | 35.5 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 8,204 | 600 | 20.6 | 13.8 | 59.4 | 22.6 | 4.2 | 26.8 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 27,642 | 594 | 18.7 | 18.0 | 66.1 | 14.6 | 1.3 | 15.9 |

Table 6.C.5 *(continuation three)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 815 | 596 | 21.1 | 19.4 | 58.5 | 18.7 | 3.4 | 22.1 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,604 | 588 | 18.4 | 28.4 | 61.4 | 9.2 | 1.0 | 10.2 |
| White (Primary ethnicity—not economically disadvantaged) | 9,612 | 609 | 21.8 | 8.1 | 46.8 | 32.9 | 12.2 | 45.1 |
| White (Primary ethnicity—economically disadvantaged) | 3,824 | 600 | 20.6 | 14.6 | 58.6 | 22.7 | 4.1 | 26.8 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 1,448 | 609 | 22.2 | 8.8 | 43.9 | 35.0 | 12.3 | 47.3 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 800 | 599 | 20.6 | 15.8 | 59.0 | 21.9 | 3.4 | 25.3 |

Table 6.C.6 Percent of Students in Each Achievement Level for Total Scores by Demographic Variables for High School

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | 107,784 | 601 | 21.8 | 14.1 | 56.6 | 22.9 | 6.4 | 29.3 |
| Male | 53,410 | 601 | 22.7 | 16.0 | 53.9 | 22.8 | 7.3 | 30.1 |
| Female | 54,332 | 601 | 20.8 | 12.1 | 59.4 | 22.9 | 5.6 | 28.5 |
| Nonbinary | 42 | 606 | 24.9 | 11.9 | 42.9 | 33.3 | 11.9 | 45.2 |
| EL | 8,644 | 581 | 13.4 | 38.6 | 59.8 | 1.5 | 0.0 | 1.6 |
| English only | 56,604 | 604 | 22.1 | 12.3 | 53.2 | 26.5 | 7.9 | 34.4 |
| RFEP | 37,345 | 600 | 20.0 | 11.9 | 62.7 | 21.1 | 4.3 | 25.4 |
| IFEP | 5,104 | 612 | 22.4 | 7.2 | 44.5 | 31.4 | 16.9 | 48.3 |
| ADEL | 69 | 586 | 11.0 | 17.4 | 82.6 | 0.0 | 0.0 | 0.0 |
| To be determined | 3 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 15 | 599 | 24.7 | 13.3 | 53.3 | 20.0 | 13.3 | 33.3 |
| Economically disadvantaged | 61,808 | 596 | 19.9 | 17.4 | 62.9 | 17.2 | 2.5 | 19.7 |
| Not economically disadvantaged | 45,976 | 608 | 22.4 | 9.6 | 48.2 | 30.5 | 11.8 | 42.2 |

Table 6.C.6 *(continuation one)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| American Indian or Alaska Native (All) | 647 | 597 | 20.4 | 17.2 | 60.6 | 18.9 | 3.4 | 22.3 |
| Asian (All) | 10,113 | 616 | 22.3 | 5.8 | 36.7 | 35.3 | 22.3 | 57.5 |
| Native Hawaiian or Other Pacific Islander (All) | 422 | 598 | 20.2 | 16.1 | 63.0 | 18.0 | 2.8 | 20.9 |
| Filipino (All) | 3,254 | 610 | 20.3 | 6.0 | 49.4 | 34.8 | 9.8 | 44.7 |
| Hispanic or Latino (All) | 58,342 | 596 | 19.5 | 17.3 | 64.0 | 16.7 | 2.1 | 18.8 |
| Black or African American (All) | 3,874 | 592 | 20.3 | 24.6 | 59.3 | 13.9 | 2.1 | 16.1 |
| White (All) | 26,977 | 607 | 22.0 | 10.1 | 49.9 | 30.4 | 9.7 | 40.1 |
| Two or more races (All) | 4,155 | 606 | 22.5 | 11.1 | 48.4 | 30.4 | 10.1 | 40.5 |
| Special education services | 9,610 | 584 | 17.3 | 36.3 | 56.7 | 6.2 | 0.8 | 7.0 |
| No special education services | 98,174 | 603 | 21.5 | 11.9 | 56.6 | 24.5 | 7.0 | 31.5 |
| Migrant education | 753 | 591 | 17.9 | 21.8 | 66.8 | 11.0 | 0.4 | 11.4 |
| Not migrant education | 107,031 | 601 | 21.8 | 14.0 | 56.6 | 22.9 | 6.5 | 29.4 |
| Military | 1,628 | 603 | 22.6 | 13.1 | 53.3 | 25.5 | 8.1 | 33.6 |
| Not military | 106,156 | 601 | 21.8 | 14.1 | 56.7 | 22.8 | 6.4 | 29.2 |
| Homeless | 3,551 | 592 | 19.2 | 21.5 | 63.5 | 13.7 | 1.3 | 15.0 |
| Not homeless | 104,233 | 601 | 21.8 | 13.8 | 56.4 | 23.2 | 6.6 | 29.8 |

Table 6.C.6 *(continuation two)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 264 | 600 | 20.1 | 13.6 | 59.8 | 22.0 | 4.5 | 26.5 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 383 | 595 | 20.3 | 19.6 | 61.1 | 16.7 | 2.6 | 19.3 |
| Asian (Primary ethnicity—not economically disadvantaged) | 6,240 | 621 | 21.2 | 3.9 | 28.8 | 38.0 | 29.2 | 67.3 |
| Asian (Primary ethnicity—economically disadvantaged) | 3,873 | 608 | 21.9 | 8.7 | 49.5 | 30.8 | 11.0 | 41.8 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 196 | 602 | 20.4 | 11.2 | 64.8 | 18.4 | 5.6 | 24.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 226 | 595 | 19.4 | 20.4 | 61.5 | 17.7 | 0.4 | 18.1 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 2,078 | 612 | 19.9 | 4.6 | 46.3 | 37.2 | 11.9 | 49.1 |
| Filipino (Primary ethnicity—economically disadvantaged) | 1,176 | 605 | 20.4 | 8.4 | 54.8 | 30.7 | 6.1 | 36.8 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 13,995 | 600 | 20.8 | 13.9 | 59.2 | 22.5 | 4.4 | 26.8 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 44,347 | 594 | 18.9 | 18.3 | 65.4 | 14.8 | 1.4 | 16.3 |

Table 6.C.6 *(continuation three)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 1,361 | 597 | 21.7 | 19.3 | 56.2 | 20.2 | 4.3 | 24.5 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 2,513 | 589 | 18.9 | 27.4 | 61.0 | 10.5 | 1.0 | 11.5 |
| White (Primary ethnicity—not economically disadvantaged) | 19,161 | 609 | 21.8 | 8.2 | 46.5 | 33.4 | 12.0 | 45.4 |
| White (Primary ethnicity—economically disadvantaged) | 7,816 | 600 | 20.8 | 14.7 | 58.2 | 23.1 | 4.0 | 27.0 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,681 | 610 | 22.3 | 8.2 | 43.4 | 34.9 | 13.4 | 48.4 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,474 | 599 | 21.0 | 16.4 | 57.5 | 22.2 | 3.9 | 26.1 |

### Appendix 6.D: Demographic Summary of Domain Achievement Levels

**Note:** Student results are suppressed and indicted as “N/A” in table 6.D.1 through table 6.D.18 where fewer than 11 students are reported in a category.

Table 6.D.1 Percent of Students in Each Achievement Level by Demographic Variables for Grade Five—Life Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 61,684 | 44.7 | 44.1 | 11.1 |
| Male | 31,297 | 45.7 | 43.4 | 11.0 |
| Female | 30,385 | 43.8 | 44.9 | 11.3 |
| Nonbinary | 2 | N/A | N/A | N/A |
| EL | 11,199 | 80.0 | 19.0 | 0.9 |
| English only | 38,568 | 39.2 | 48.1 | 12.7 |
| RFEP | 9,519 | 32.9 | 54.9 | 12.2 |
| IFEP | 2,358 | 16.0 | 55.0 | 29.0 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 11 | 54.5 | 36.4 | 9.1 |
| English proficiency unknown | 29 | 44.8 | 41.4 | 13.8 |
| Economically disadvantaged | 33,647 | 59.1 | 36.0 | 4.9 |
| Not economically disadvantaged | 28,037 | 27.4 | 53.9 | 18.6 |
| American Indian or Alaska Native (All) | 497 | 54.9 | 39.4 | 5.6 |
| Asian (All) | 7,102 | 20.2 | 53.1 | 26.7 |
| Native Hawaiian or Other Pacific Islander (All) | 232 | 55.6 | 37.5 | 6.9 |
| Filipino (All) | 1,126 | 31.8 | 53.6 | 14.7 |
| Hispanic or Latino (All) | 29,828 | 58.7 | 36.5 | 4.8 |
| Black or African American (All) | 2,220 | 63.4 | 32.2 | 4.4 |
| White (All) | 17,101 | 31.1 | 53.3 | 15.5 |
| Two or more races (All) | 3,578 | 31.9 | 51.9 | 16.2 |
| Special education services | 6,964 | 72.7 | 23.2 | 4.1 |
| No special education services | 54,720 | 41.2 | 46.8 | 12.0 |

Table 6.D.1 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Migrant education | 621 | 72.3 | 24.6 | 3.1 |
| Not migrant education | 61,063 | 44.4 | 44.3 | 11.2 |
| Military | 684 | 39.9 | 49.6 | 10.5 |
| Not military | 61,000 | 44.8 | 44.1 | 11.1 |
| Homeless | 1,442 | 65.8 | 31.4 | 2.8 |
| Not homeless | 60,242 | 44.2 | 44.4 | 11.3 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 147 | 38.8 | 52.4 | 8.8 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 350 | 61.7 | 34.0 | 4.3 |
| Asian (Primary ethnicity—not economically disadvantaged) | 5,445 | 12.5 | 55.8 | 31.6 |
| Asian (Primary ethnicity—economically disadvantaged) | 1,657 | 45.6 | 43.9 | 10.4 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 100 | 40.0 | 47.0 | 13.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 132 | 67.4 | 30.3 | 2.3 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 783 | 27.7 | 56.4 | 15.8 |
| Filipino (Primary ethnicity—economically disadvantaged) | 343 | 41.1 | 46.9 | 12.0 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 7,273 | 43.3 | 47.8 | 8.9 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 22,555 | 63.7 | 32.8 | 3.5 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 679 | 48.3 | 44.5 | 7.2 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,541 | 70.1 | 26.8 | 3.1 |
| White (Primary ethnicity—not economically disadvantaged) | 11,320 | 23.9 | 56.9 | 19.2 |
| White (Primary ethnicity—economically disadvantaged) | 5,781 | 45.4 | 46.2 | 8.4 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,290 | 22.7 | 56.4 | 21.0 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,288 | 48.2 | 43.9 | 7.8 |

Table 6.D.2 Percent of Students in Each Achievement Level by Demographic Variables for Grade Eight—Life Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 72,308 | 40.7 | 50.5 | 8.8 |
| Male | 37,005 | 43.7 | 47.3 | 9.0 |
| Female | 35,280 | 37.5 | 53.8 | 8.7 |
| Nonbinary | 23 | 39.1 | 56.5 | 4.3 |
| EL | 9,916 | 80.7 | 19.2 | 0.2 |
| English only | 41,524 | 36.0 | 53.8 | 10.1 |
| RFEP | 17,583 | 33.8 | 59.2 | 7.1 |
| IFEP | 3,249 | 15.4 | 56.1 | 28.5 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 6 | N/A | N/A | N/A |
| English proficiency unknown | 30 | 36.7 | 60.0 | 3.3 |
| Economically disadvantaged | 39,483 | 53.1 | 43.9 | 3.0 |
| Not economically disadvantaged | 32,825 | 25.8 | 58.5 | 15.8 |
| American Indian or Alaska Native (All) | 532 | 56.6 | 40.0 | 3.4 |
| Asian (All) | 7,733 | 14.5 | 56.5 | 29.0 |
| Native Hawaiian or Other Pacific Islander (All) | 257 | 45.5 | 51.8 | 2.7 |
| Filipino (All) | 1,767 | 23.3 | 63.8 | 12.9 |
| Hispanic or Latino (All) | 36,312 | 52.5 | 44.6 | 2.9 |
| Black or African American (All) | 2,594 | 59.4 | 37.8 | 2.8 |
| White (All) | 19,397 | 29.8 | 58.7 | 11.6 |
| Two or more races (All) | 3,716 | 29.8 | 56.4 | 13.8 |
| Special education services | 7,865 | 75.9 | 22.6 | 1.5 |
| No special education services | 64,443 | 36.4 | 53.9 | 9.7 |
| Migrant education | 640 | 61.1 | 37.8 | 1.1 |
| Not migrant education | 71,668 | 40.5 | 50.6 | 8.9 |
| Military | 1,030 | 31.0 | 58.2 | 10.9 |
| Not military | 71,278 | 40.8 | 50.4 | 8.8 |

Table 6.D.2 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 1,663 | 60.6 | 37.8 | 1.6 |
| Not homeless | 70,645 | 40.2 | 50.8 | 9.0 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 179 | 47.5 | 48.0 | 4.5 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 353 | 61.2 | 36.0 | 2.8 |
| Asian (Primary ethnicity—not economically disadvantaged) | 5,776 | 8.5 | 56.5 | 35.0 |
| Asian (Primary ethnicity—economically disadvantaged) | 1,957 | 32.4 | 56.5 | 11.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 116 | 37.9 | 57.8 | 4.3 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 141 | 51.8 | 46.8 | 1.4 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 1,212 | 20.1 | 64.8 | 15.1 |
| Filipino (Primary ethnicity—economically disadvantaged) | 555 | 30.1 | 61.8 | 8.1 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 8,990 | 40.3 | 53.7 | 5.9 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 27,322 | 56.5 | 41.6 | 1.9 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 838 | 44.7 | 50.1 | 5.1 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,756 | 66.4 | 31.9 | 1.7 |
| White (Primary ethnicity—not economically disadvantaged) | 13,277 | 23.3 | 62.0 | 14.6 |
| White (Primary ethnicity—economically disadvantaged) | 6,120 | 43.7 | 51.4 | 5.0 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,437 | 20.4 | 61.4 | 18.2 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,279 | 47.8 | 46.7 | 5.6 |

Table 6.D.3 Percent of Students in Each Achievement Level by Demographic Variables for Grade Ten—Life Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 2,229 | 28.1 | 56.6 | 15.3 |
| Male | 1,112 | 29.6 | 54.9 | 15.6 |
| Female | 1,116 | 26.7 | 58.4 | 14.9 |
| Nonbinary | 1 | N/A | N/A | N/A |
| EL | 153 | 73.2 | 26.8 | 0.0 |
| English only | 1,425 | 22.9 | 61.2 | 15.9 |
| RFEP | 510 | 34.1 | 55.1 | 10.8 |
| IFEP | 141 | 9.9 | 48.2 | 41.8 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 0 | N/A | N/A | N/A |
| English proficiency unknown | 0 | N/A | N/A | N/A |
| Economically disadvantaged | 982 | 42.7 | 53.7 | 3.7 |
| Not economically disadvantaged | 1,247 | 16.7 | 58.9 | 24.4 |
| American Indian or Alaska Native (All) | 22 | 40.9 | 59.1 | 0.0 |
| Asian (All) | 304 | 5.9 | 40.1 | 53.9 |
| Native Hawaiian or Other Pacific Islander (All) | 7 | N/A | N/A | N/A |
| Filipino (All) | 19 | 15.8 | 68.4 | 15.8 |
| Hispanic or Latino (All) | 880 | 41.9 | 55.6 | 2.5 |
| Black or African American (All) | 52 | 38.5 | 53.8 | 7.7 |
| White (All) | 821 | 22.0 | 64.2 | 13.8 |
| Two or more races (All) | 124 | 19.4 | 54.0 | 26.6 |
| Special education services | 220 | 56.4 | 41.8 | 1.8 |
| No special education services | 2,009 | 25.0 | 58.2 | 16.7 |
| Migrant education | 13 | 76.9 | 23.1 | 0.0 |
| Not migrant education | 2,216 | 27.8 | 56.8 | 15.3 |
| Military | 14 | 28.6 | 64.3 | 7.1 |
| Not military | 2,215 | 28.1 | 56.6 | 15.3 |

Table 6.D.3 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 28 | 60.7 | 39.3 | 0.0 |
| Not homeless | 2,201 | 27.7 | 56.8 | 15.4 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 9 | N/A | N/A | N/A |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 13 | 46.2 | 53.8 | 0.0 |
| Asian (Primary ethnicity—not economically disadvantaged) | 276 | 3.6 | 39.9 | 56.5 |
| Asian (Primary ethnicity—economically disadvantaged) | 28 | 28.6 | 42.9 | 28.6 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 6 | N/A | N/A | N/A |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1 | N/A | N/A | N/A |
| Filipino (Primary ethnicity—not economically disadvantaged) | 16 | 6.3 | 75.0 | 18.8 |
| Filipino (Primary ethnicity—economically disadvantaged) | 3 | N/A | N/A | N/A |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 272 | 29.8 | 65.4 | 4.8 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 608 | 47.4 | 51.2 | 1.5 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 26 | 26.9 | 61.5 | 11.5 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 26 | 50.0 | 46.2 | 3.8 |
| White (Primary ethnicity—not economically disadvantaged) | 557 | 16.5 | 65.9 | 17.6 |
| White (Primary ethnicity—economically disadvantaged) | 264 | 33.7 | 60.6 | 5.7 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 85 | 14.1 | 50.6 | 35.3 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 39 | 30.8 | 61.5 | 7.7 |

Table 6.D.4 Percent of Students in Each Achievement Level by Demographic Variables for Grade Eleven—Life Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 42,955 | 32.0 | 57.7 | 10.3 |
| Male | 21,412 | 34.7 | 54.8 | 10.5 |
| Female | 21,525 | 29.4 | 60.6 | 10.1 |
| Nonbinary | 18 | 11.1 | 72.2 | 16.7 |
| EL | 3,284 | 72.1 | 27.8 | 0.1 |
| English only | 23,820 | 28.9 | 59.0 | 12.1 |
| RFEP | 14,002 | 30.1 | 62.1 | 7.8 |
| IFEP | 1,840 | 16.3 | 59.9 | 23.8 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 1 | N/A | N/A | N/A |
| English proficiency unknown | 8 | N/A | N/A | N/A |
| Economically disadvantaged | 23,177 | 39.0 | 56.0 | 5.0 |
| Not economically disadvantaged | 19,778 | 23.9 | 59.6 | 16.5 |
| American Indian or Alaska Native (All) | 302 | 40.7 | 54.6 | 4.6 |
| Asian (All) | 4,444 | 15.0 | 57.7 | 27.3 |
| Native Hawaiian or Other Pacific Islander (All) | 169 | 35.5 | 58.6 | 5.9 |
| Filipino (All) | 1,300 | 16.2 | 66.9 | 16.9 |
| Hispanic or Latino (All) | 21,042 | 39.7 | 56.0 | 4.3 |
| Black or African American (All) | 1,350 | 45.6 | 50.1 | 4.3 |
| White (All) | 12,586 | 26.1 | 60.1 | 13.8 |
| Two or more races (All) | 1,762 | 25.3 | 60.2 | 14.6 |
| Special education services | 3,615 | 66.0 | 32.5 | 1.5 |
| No special education services | 39,340 | 28.9 | 60.0 | 11.1 |
| Migrant education | 264 | 46.6 | 51.9 | 1.5 |
| Not migrant education | 42,691 | 31.9 | 57.7 | 10.3 |
| Military | 641 | 30.0 | 55.1 | 15.0 |
| Not military | 42,314 | 32.1 | 57.7 | 10.2 |

Table 6.D.4 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 1,318 | 47.9 | 49.0 | 3.1 |
| Not homeless | 41,637 | 31.5 | 58.0 | 10.5 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 137 | 32.8 | 62.8 | 4.4 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 165 | 47.3 | 47.9 | 4.8 |
| Asian (Primary ethnicity—not economically disadvantaged) | 2,727 | 10.5 | 54.2 | 35.3 |
| Asian (Primary ethnicity—economically disadvantaged) | 1,717 | 22.1 | 63.4 | 14.5 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 81 | 32.1 | 58.0 | 9.9 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 88 | 38.6 | 59.1 | 2.3 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 859 | 14.0 | 66.7 | 19.3 |
| Filipino (Primary ethnicity—economically disadvantaged) | 441 | 20.4 | 67.3 | 12.2 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 5,413 | 33.5 | 59.0 | 7.5 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 15,629 | 41.8 | 54.9 | 3.2 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 504 | 38.5 | 55.2 | 6.3 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 846 | 49.8 | 47.2 | 3.1 |
| White (Primary ethnicity—not economically disadvantaged) | 8,923 | 22.5 | 61.0 | 16.5 |
| White (Primary ethnicity—economically disadvantaged) | 3,663 | 34.9 | 58.1 | 7.0 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 1,134 | 19.7 | 61.6 | 18.7 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 628 | 35.4 | 57.5 | 7.2 |

Table 6.D.5 Percent of Students in Each Achievement Level by Demographic Variables for Grade Twelve—Life Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 60,651 | 34.6 | 57.0 | 8.4 |
| Male | 29,874 | 36.6 | 54.6 | 8.8 |
| Female | 30,755 | 32.7 | 59.3 | 8.0 |
| Nonbinary | 22 | 27.3 | 54.5 | 18.2 |
| EL | 4,926 | 72.4 | 27.5 | 0.1 |
| English only | 30,582 | 30.9 | 58.6 | 10.6 |
| RFEP | 22,014 | 33.2 | 60.9 | 5.9 |
| IFEP | 3,052 | 20.6 | 61.4 | 18.0 |
| ADEL | 68 | 64.7 | 35.3 | 0.0 |
| To be determined | 2 | N/A | N/A | N/A |
| English proficiency unknown | 7 | N/A | N/A | N/A |
| Economically disadvantaged | 36,194 | 40.9 | 55.1 | 4.0 |
| Not economically disadvantaged | 24,457 | 25.4 | 59.8 | 14.8 |
| American Indian or Alaska Native (All) | 316 | 40.5 | 54.4 | 5.1 |
| Asian (All) | 5,261 | 17.2 | 58.7 | 24.0 |
| Native Hawaiian or Other Pacific Islander (All) | 233 | 37.3 | 59.7 | 3.0 |
| Filipino (All) | 1,889 | 21.2 | 67.0 | 11.8 |
| Hispanic or Latino (All) | 35,064 | 40.9 | 55.2 | 3.9 |
| Black or African American (All) | 2,350 | 49.3 | 47.6 | 3.1 |
| White (All) | 13,315 | 25.4 | 60.8 | 13.8 |
| Two or more races (All) | 2,223 | 27.4 | 59.3 | 13.2 |
| Special education services | 5,505 | 63.4 | 35.3 | 1.3 |
| No special education services | 55,146 | 31.8 | 59.2 | 9.1 |
| Migrant education | 462 | 53.0 | 45.2 | 1.7 |
| Not migrant education | 60,189 | 34.5 | 57.1 | 8.4 |
| Military | 947 | 34.4 | 57.7 | 7.9 |
| Not military | 59,704 | 34.6 | 57.0 | 8.4 |

Table 6.D.5 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 2,090 | 47.4 | 50.0 | 2.6 |
| Not homeless | 58,561 | 34.2 | 57.2 | 8.6 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 114 | 33.3 | 61.4 | 5.3 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 202 | 44.6 | 50.5 | 5.0 |
| Asian (Primary ethnicity—not economically disadvantaged) | 3,188 | 12.7 | 56.7 | 30.6 |
| Asian (Primary ethnicity—economically disadvantaged) | 2,073 | 24.2 | 61.8 | 14.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 105 | 33.3 | 61.9 | 4.8 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 128 | 40.6 | 57.8 | 1.6 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 1,188 | 19.3 | 67.1 | 13.6 |
| Filipino (Primary ethnicity—economically disadvantaged) | 701 | 24.5 | 66.8 | 8.7 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 8,087 | 34.1 | 58.8 | 7.0 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 26,977 | 42.9 | 54.2 | 3.0 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 795 | 40.4 | 54.1 | 5.5 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,555 | 53.9 | 44.2 | 1.9 |
| White (Primary ethnicity—not economically disadvantaged) | 9,543 | 21.9 | 61.2 | 16.9 |
| White (Primary ethnicity—economically disadvantaged) | 3,772 | 34.1 | 59.8 | 6.1 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 1,437 | 22.5 | 59.9 | 17.5 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 786 | 36.4 | 58.3 | 5.3 |

Table 6.D.6 Percent of Students in Each Achievement Level by Demographic Variables for High School—Life Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 105,835 | 33.4 | 57.3 | 9.3 |
| Male | 52,398 | 35.7 | 54.7 | 9.6 |
| Female | 53,396 | 31.2 | 59.8 | 9.0 |
| Nonbinary | 41 | 19.5 | 61.0 | 19.5 |
| EL | 8,363 | 72.3 | 27.6 | 0.1 |
| English only | 55,827 | 29.8 | 58.8 | 11.4 |
| RFEP | 36,526 | 32.0 | 61.3 | 6.7 |
| IFEP | 5,033 | 18.8 | 60.5 | 20.7 |
| ADEL | 68 | 64.7 | 35.3 | 0.0 |
| To be determined | 3 | N/A | N/A | N/A |
| English proficiency unknown | 15 | 46.7 | 33.3 | 20.0 |
| Economically disadvantaged | 60,353 | 40.2 | 55.4 | 4.4 |
| Not economically disadvantaged | 45,482 | 24.5 | 59.7 | 15.8 |
| American Indian or Alaska Native (All) | 640 | 40.6 | 54.7 | 4.7 |
| Asian (All) | 10,009 | 15.9 | 57.7 | 26.4 |
| Native Hawaiian or Other Pacific Islander (All) | 409 | 36.7 | 58.9 | 4.4 |
| Filipino (All) | 3,208 | 19.1 | 67.0 | 13.9 |
| Hispanic or Latino (All) | 56,986 | 40.5 | 55.5 | 4.0 |
| Black or African American (All) | 3,752 | 47.8 | 48.6 | 3.6 |
| White (All) | 26,722 | 25.6 | 60.6 | 13.8 |
| Two or more races (All) | 4,109 | 26.3 | 59.5 | 14.2 |
| Special education services | 9,340 | 64.3 | 34.3 | 1.4 |
| No special education services | 96,495 | 30.5 | 59.5 | 10.1 |
| Migrant education | 739 | 51.2 | 47.2 | 1.6 |
| Not migrant education | 105,096 | 33.3 | 57.3 | 9.3 |
| Military | 1,602 | 32.6 | 56.7 | 10.7 |
| Not military | 104,233 | 33.4 | 57.3 | 9.3 |

Table 6.D.6 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 3,436 | 47.7 | 49.5 | 2.8 |
| Not homeless | 102,399 | 33.0 | 57.5 | 9.5 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 260 | 33.1 | 62.3 | 4.6 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 380 | 45.8 | 49.5 | 4.7 |
| Asian (Primary ethnicity—not economically disadvantaged) | 6,191 | 11.4 | 54.8 | 33.8 |
| Asian (Primary ethnicity—economically disadvantaged) | 3,818 | 23.3 | 62.4 | 14.3 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 192 | 32.8 | 59.9 | 7.3 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 217 | 40.1 | 58.1 | 1.8 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 2,063 | 17.0 | 67.0 | 16.0 |
| Filipino (Primary ethnicity—economically disadvantaged) | 1,145 | 23.1 | 66.9 | 10.0 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 13,772 | 33.8 | 59.0 | 7.2 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 43,214 | 42.6 | 54.4 | 3.0 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 1,325 | 39.4 | 54.6 | 6.0 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 2,427 | 52.4 | 45.3 | 2.3 |
| White (Primary ethnicity—not economically disadvantaged) | 19,023 | 22.0 | 61.2 | 16.7 |
| White (Primary ethnicity—economically disadvantaged) | 7,699 | 34.4 | 59.0 | 6.5 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,656 | 21.0 | 60.4 | 18.6 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,453 | 35.8 | 58.0 | 6.2 |

Table 6.D.7 Percent of Students in Each Achievement Level by Demographic Variables for Grade Five—Physical Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 61,763 | 47.7 | 40.8 | 11.4 |
| Male | 31,341 | 49.0 | 38.9 | 12.1 |
| Female | 30,420 | 46.4 | 42.8 | 10.7 |
| Nonbinary | 2 | N/A | N/A | N/A |
| EL | 11,222 | 78.6 | 20.2 | 1.2 |
| English only | 38,605 | 43.2 | 44.1 | 12.7 |
| RFEP | 9,534 | 37.4 | 50.0 | 12.7 |
| IFEP | 2,362 | 17.7 | 48.1 | 34.2 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 11 | 54.5 | 18.2 | 27.3 |
| English proficiency unknown | 29 | 48.3 | 37.9 | 13.8 |
| Economically disadvantaged | 33,716 | 61.8 | 33.6 | 4.5 |
| Not economically disadvantaged | 28,047 | 30.8 | 49.5 | 19.7 |
| American Indian or Alaska Native (All) | 498 | 59.2 | 36.3 | 4.4 |
| Asian (All) | 7,102 | 20.8 | 47.6 | 31.5 |
| Native Hawaiian or Other Pacific Islander (All) | 232 | 58.6 | 36.2 | 5.2 |
| Filipino (All) | 1,126 | 33.5 | 52.1 | 14.4 |
| Hispanic or Latino (All) | 29,892 | 61.6 | 33.8 | 4.5 |
| Black or African American (All) | 2,222 | 65.6 | 30.0 | 4.4 |
| White (All) | 17,110 | 35.1 | 49.9 | 15.0 |
| Two or more races (All) | 3,581 | 36.7 | 46.6 | 16.7 |
| Special education services | 6,973 | 73.9 | 21.6 | 4.5 |
| No special education services | 54,790 | 44.4 | 43.3 | 12.3 |
| Migrant education | 622 | 72.7 | 24.9 | 2.4 |
| Not migrant education | 61,141 | 47.5 | 41.0 | 11.5 |
| Military | 684 | 41.4 | 47.5 | 11.1 |
| Not military | 61,079 | 47.8 | 40.8 | 11.4 |

Table 6.D.7 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 1,446 | 69.0 | 27.8 | 3.2 |
| Not homeless | 60,317 | 47.2 | 41.1 | 11.6 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 147 | 49.0 | 45.6 | 5.4 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 351 | 63.5 | 32.5 | 4.0 |
| Asian (Primary ethnicity—not economically disadvantaged) | 5,445 | 13.1 | 49.4 | 37.5 |
| Asian (Primary ethnicity—economically disadvantaged) | 1,657 | 46.2 | 41.9 | 11.9 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 100 | 48.0 | 44.0 | 8.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 132 | 66.7 | 30.3 | 3.0 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 783 | 29.6 | 53.8 | 16.6 |
| Filipino (Primary ethnicity—economically disadvantaged) | 343 | 42.3 | 48.4 | 9.3 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 7,278 | 47.8 | 43.3 | 8.8 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 22,614 | 66.1 | 30.8 | 3.1 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 680 | 52.1 | 39.3 | 8.7 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,542 | 71.6 | 25.9 | 2.5 |
| White (Primary ethnicity—not economically disadvantaged) | 11,321 | 27.6 | 53.6 | 18.8 |
| White (Primary ethnicity—economically disadvantaged) | 5,789 | 49.7 | 42.7 | 7.6 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,293 | 27.1 | 51.2 | 21.7 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,288 | 54.0 | 38.2 | 7.8 |

Table 6.D.8 Percent of Students in Each Achievement Level by Demographic Variables for Grade Eight—Physical Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 72,654 | 40.1 | 50.2 | 9.7 |
| Male | 37,176 | 41.5 | 47.4 | 11.1 |
| Female | 35,455 | 38.7 | 53.1 | 8.2 |
| Nonbinary | 23 | 30.4 | 56.5 | 13.0 |
| EL | 9,980 | 78.6 | 21.0 | 0.3 |
| English only | 41,666 | 35.2 | 53.5 | 11.3 |
| RFEP | 17,714 | 34.6 | 57.6 | 7.8 |
| IFEP | 3,258 | 15.3 | 56.4 | 28.3 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 6 | N/A | N/A | N/A |
| English proficiency unknown | 30 | 46.7 | 43.3 | 10.0 |
| Economically disadvantaged | 39,750 | 52.4 | 44.2 | 3.5 |
| Not economically disadvantaged | 32,904 | 25.3 | 57.4 | 17.3 |
| American Indian or Alaska Native (All) | 539 | 53.2 | 43.2 | 3.5 |
| Asian (All) | 7,748 | 14.5 | 56.1 | 29.4 |
| Native Hawaiian or Other Pacific Islander (All) | 258 | 48.1 | 49.2 | 2.7 |
| Filipino (All) | 1,770 | 23.6 | 64.0 | 12.4 |
| Hispanic or Latino (All) | 36,567 | 52.0 | 44.7 | 3.3 |
| Black or African American (All) | 2,612 | 56.4 | 41.2 | 2.4 |
| White (All) | 19,436 | 28.9 | 57.3 | 13.8 |
| Two or more races (All) | 3,724 | 28.9 | 55.8 | 15.4 |
| Special education services | 7,906 | 74.6 | 23.6 | 1.8 |
| No special education services | 64,748 | 35.9 | 53.4 | 10.7 |
| Migrant education | 643 | 61.9 | 35.9 | 2.2 |
| Not migrant education | 72,011 | 39.9 | 50.3 | 9.8 |
| Military | 1,031 | 31.4 | 57.7 | 10.9 |
| Not military | 71,623 | 40.2 | 50.1 | 9.7 |

Table 6.D.8 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 1,679 | 59.2 | 38.2 | 2.6 |
| Not homeless | 70,975 | 39.7 | 50.5 | 9.9 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 180 | 46.1 | 48.3 | 5.6 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 359 | 56.8 | 40.7 | 2.5 |
| Asian (Primary ethnicity—not economically disadvantaged) | 5,782 | 8.4 | 56.4 | 35.2 |
| Asian (Primary ethnicity—economically disadvantaged) | 1,966 | 32.5 | 55.2 | 12.3 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 116 | 38.8 | 56.9 | 4.3 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 142 | 55.6 | 43.0 | 1.4 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 1,213 | 20.7 | 64.6 | 14.8 |
| Filipino (Primary ethnicity—economically disadvantaged) | 557 | 29.8 | 62.8 | 7.4 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 9,030 | 40.0 | 53.2 | 6.8 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 27,537 | 56.0 | 41.8 | 2.2 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 846 | 40.9 | 54.6 | 4.5 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,766 | 63.8 | 34.8 | 1.4 |
| White (Primary ethnicity—not economically disadvantaged) | 13,296 | 22.7 | 60.0 | 17.3 |
| White (Primary ethnicity—economically disadvantaged) | 6,140 | 42.3 | 51.3 | 6.3 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,441 | 20.1 | 59.4 | 20.5 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,283 | 45.5 | 48.9 | 5.6 |

Table 6.D.9 Percent of Students in Each Achievement Level by Demographic Variables for Grade Ten—Physical Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 2,232 | 36.1 | 51.4 | 12.5 |
| Male | 1,115 | 37.4 | 48.9 | 13.7 |
| Female | 1,116 | 34.8 | 53.9 | 11.3 |
| Nonbinary | 1 | N/A | N/A | N/A |
| EL | 153 | 77.1 | 22.9 | 0.0 |
| English only | 1,425 | 32.0 | 56.0 | 12.0 |
| RFEP | 512 | 41.0 | 48.6 | 10.4 |
| IFEP | 142 | 14.8 | 45.8 | 39.4 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 0 | N/A | N/A | N/A |
| English proficiency unknown | 0 | N/A | N/A | N/A |
| Economically disadvantaged | 982 | 51.1 | 46.4 | 2.4 |
| Not economically disadvantaged | 1,250 | 24.2 | 55.3 | 20.5 |
| American Indian or Alaska Native (All) | 22 | 59.1 | 36.4 | 4.5 |
| Asian (All) | 306 | 7.8 | 45.4 | 46.7 |
| Native Hawaiian or Other Pacific Islander (All) | 7 | N/A | N/A | N/A |
| Filipino (All) | 19 | 21.1 | 73.7 | 5.3 |
| Hispanic or Latino (All) | 882 | 53.5 | 44.1 | 2.4 |
| Black or African American (All) | 52 | 48.1 | 46.2 | 5.8 |
| White (All) | 821 | 28.3 | 61.9 | 9.9 |
| Two or more races (All) | 123 | 26.8 | 49.6 | 23.6 |
| Special education services | 220 | 69.5 | 28.6 | 1.8 |
| No special education services | 2,012 | 32.4 | 53.9 | 13.7 |
| Migrant education | 13 | 84.6 | 15.4 | 0.0 |
| Not migrant education | 2,219 | 35.8 | 51.6 | 12.6 |
| Military | 14 | 35.7 | 57.1 | 7.1 |
| Not military | 2,218 | 36.1 | 51.4 | 12.6 |

Table 6.D.9 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 28 | 67.9 | 32.1 | 0.0 |
| Not homeless | 2,204 | 35.7 | 51.6 | 12.7 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 9 | N/A | N/A | N/A |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 13 | 69.2 | 30.8 | 0.0 |
| Asian (Primary ethnicity—not economically disadvantaged) | 278 | 4.7 | 46.0 | 49.3 |
| Asian (Primary ethnicity—economically disadvantaged) | 28 | 39.3 | 39.3 | 21.4 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 6 | N/A | N/A | N/A |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1 | N/A | N/A | N/A |
| Filipino (Primary ethnicity—not economically disadvantaged) | 16 | 18.8 | 75.0 | 6.3 |
| Filipino (Primary ethnicity—economically disadvantaged) | 3 | N/A | N/A | N/A |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 273 | 47.3 | 47.3 | 5.5 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 609 | 56.3 | 42.7 | 1.0 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 26 | 30.8 | 57.7 | 11.5 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 26 | 65.4 | 34.6 | 0.0 |
| White (Primary ethnicity—not economically disadvantaged) | 557 | 23.0 | 64.3 | 12.7 |
| White (Primary ethnicity—economically disadvantaged) | 264 | 39.4 | 56.8 | 3.8 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 85 | 20.0 | 48.2 | 31.8 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 38 | 42.1 | 52.6 | 5.3 |

Table 6.D.10 Percent of Students in Each Achievement Level by Demographic Variables for Grade Eleven—Physical Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 42,982 | 40.8 | 49.9 | 9.3 |
| Male | 21,431 | 42.3 | 47.7 | 10.1 |
| Female | 21,533 | 39.4 | 52.1 | 8.4 |
| Nonbinary | 18 | 38.9 | 44.4 | 16.7 |
| EL | 3,286 | 78.4 | 21.3 | 0.3 |
| English only | 23,831 | 37.8 | 52.1 | 10.1 |
| RFEP | 14,015 | 39.4 | 52.4 | 8.1 |
| IFEP | 1,841 | 23.9 | 53.6 | 22.5 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 1 | N/A | N/A | N/A |
| English proficiency unknown | 8 | N/A | N/A | N/A |
| Economically disadvantaged | 23,197 | 48.8 | 46.4 | 4.8 |
| Not economically disadvantaged | 19,785 | 31.6 | 54.0 | 14.5 |
| American Indian or Alaska Native (All) | 302 | 53.3 | 41.7 | 5.0 |
| Asian (All) | 4,445 | 21.0 | 53.0 | 26.1 |
| Native Hawaiian or Other Pacific Islander (All) | 169 | 43.2 | 51.5 | 5.3 |
| Filipino (All) | 1,300 | 24.2 | 58.5 | 17.2 |
| Hispanic or Latino (All) | 21,065 | 49.0 | 46.7 | 4.3 |
| Black or African American (All) | 1,354 | 56.5 | 38.9 | 4.6 |
| White (All) | 12,584 | 35.0 | 54.0 | 11.0 |
| Two or more races (All) | 1,763 | 33.2 | 54.5 | 12.3 |
| Special education services | 3,622 | 73.7 | 25.2 | 1.1 |
| No special education services | 39,360 | 37.8 | 52.2 | 10.0 |
| Migrant education | 264 | 59.5 | 37.9 | 2.7 |
| Not migrant education | 42,718 | 40.7 | 50.0 | 9.3 |
| Military | 637 | 38.5 | 46.8 | 14.8 |
| Not military | 42,345 | 40.9 | 49.9 | 9.2 |

Table 6.D.10 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 1,323 | 55.0 | 41.8 | 3.2 |
| Not homeless | 41,659 | 40.4 | 50.2 | 9.5 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 137 | 47.4 | 44.5 | 8.0 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 165 | 58.2 | 39.4 | 2.4 |
| Asian (Primary ethnicity—not economically disadvantaged) | 2,727 | 15.3 | 51.4 | 33.3 |
| Asian (Primary ethnicity—economically disadvantaged) | 1,718 | 30.0 | 55.4 | 14.7 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 81 | 35.8 | 54.3 | 9.9 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 88 | 50.0 | 48.9 | 1.1 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 859 | 20.4 | 59.3 | 20.4 |
| Filipino (Primary ethnicity—economically disadvantaged) | 441 | 31.7 | 57.1 | 11.1 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 5,418 | 42.3 | 51.0 | 6.7 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 15,647 | 51.3 | 45.2 | 3.5 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 507 | 46.5 | 46.4 | 7.1 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 847 | 62.5 | 34.5 | 3.1 |
| White (Primary ethnicity—not economically disadvantaged) | 8,921 | 30.5 | 56.3 | 13.2 |
| White (Primary ethnicity—economically disadvantaged) | 3,663 | 45.9 | 48.5 | 5.5 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 1,135 | 27.5 | 56.7 | 15.8 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 628 | 43.5 | 50.5 | 6.1 |

Table 6.D.11 Percent of Students in Each Achievement Level by Demographic Variables for Grade Twelve—Physical Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 60,701 | 43.7 | 48.6 | 7.8 |
| Male | 29,909 | 44.3 | 46.8 | 9.0 |
| Female | 30,770 | 43.0 | 50.3 | 6.7 |
| Nonbinary | 22 | 40.9 | 54.5 | 4.5 |
| EL | 4,926 | 78.0 | 21.6 | 0.4 |
| English only | 30,599 | 40.1 | 50.6 | 9.3 |
| RFEP | 22,043 | 42.7 | 51.2 | 6.1 |
| IFEP | 3,056 | 30.1 | 53.3 | 16.6 |
| ADEL | 68 | 64.7 | 35.3 | 0.0 |
| To be determined | 2 | N/A | N/A | N/A |
| English proficiency unknown | 7 | N/A | N/A | N/A |
| Economically disadvantaged | 36,228 | 50.2 | 45.6 | 4.2 |
| Not economically disadvantaged | 24,473 | 33.9 | 53.0 | 13.1 |
| American Indian or Alaska Native (All) | 316 | 50.9 | 44.0 | 5.1 |
| Asian (All) | 5,258 | 23.0 | 53.5 | 23.6 |
| Native Hawaiian or Other Pacific Islander (All) | 233 | 46.4 | 50.6 | 3.0 |
| Filipino (All) | 1,891 | 30.2 | 57.7 | 12.1 |
| Hispanic or Latino (All) | 35,102 | 50.4 | 45.6 | 4.0 |
| Black or African American (All) | 2,355 | 60.3 | 36.5 | 3.2 |
| White (All) | 13,321 | 33.9 | 54.8 | 11.3 |
| Two or more races (All) | 2,225 | 36.7 | 51.6 | 11.8 |
| Special education services | 5,510 | 71.6 | 27.0 | 1.4 |
| No special education services | 55,191 | 40.9 | 50.7 | 8.4 |
| Migrant education | 462 | 60.4 | 36.4 | 3.2 |
| Not migrant education | 60,239 | 43.5 | 48.7 | 7.8 |
| Military | 950 | 42.8 | 48.5 | 8.6 |
| Not military | 59,751 | 43.7 | 48.6 | 7.8 |

Table 6.D.11 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 2,093 | 54.4 | 42.7 | 2.9 |
| Not homeless | 58,608 | 43.3 | 48.8 | 8.0 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 114 | 48.2 | 46.5 | 5.3 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 202 | 52.5 | 42.6 | 5.0 |
| Asian (Primary ethnicity—not economically disadvantaged) | 3,187 | 18.2 | 52.1 | 29.7 |
| Asian (Primary ethnicity—economically disadvantaged) | 2,071 | 30.3 | 55.5 | 14.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 105 | 41.9 | 52.4 | 5.7 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 128 | 50.0 | 49.2 | 0.8 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 1,191 | 28.2 | 57.7 | 14.1 |
| Filipino (Primary ethnicity—economically disadvantaged) | 700 | 33.7 | 57.7 | 8.6 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 8,093 | 43.7 | 49.7 | 6.6 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 27,009 | 52.4 | 44.4 | 3.2 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 797 | 50.2 | 43.9 | 5.9 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,558 | 65.4 | 32.7 | 1.9 |
| White (Primary ethnicity—not economically disadvantaged) | 9,548 | 30.5 | 56.2 | 13.3 |
| White (Primary ethnicity—economically disadvantaged) | 3,773 | 42.7 | 51.2 | 6.2 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 1,438 | 31.1 | 53.8 | 15.2 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 787 | 46.9 | 47.5 | 5.6 |

Table 6.D.12 Percent of Students in Each Achievement Level by Demographic Variables for High School—Physical Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 105,915 | 42.4 | 49.2 | 8.5 |
| Male | 52,455 | 43.3 | 47.2 | 9.5 |
| Female | 53,419 | 41.4 | 51.1 | 7.5 |
| Nonbinary | 41 | 39.0 | 48.8 | 12.2 |
| EL | 8,365 | 78.1 | 21.5 | 0.4 |
| English only | 55,855 | 38.9 | 51.3 | 9.7 |
| RFEP | 36,570 | 41.5 | 51.6 | 6.9 |
| IFEP | 5,039 | 27.4 | 53.2 | 19.4 |
| ADEL | 68 | 64.7 | 35.3 | 0.0 |
| To be determined | 3 | N/A | N/A | N/A |
| English proficiency unknown | 15 | 53.3 | 26.7 | 20.0 |
| Economically disadvantaged | 60,407 | 49.7 | 45.9 | 4.4 |
| Not economically disadvantaged | 45,508 | 32.6 | 53.5 | 13.9 |
| American Indian or Alaska Native (All) | 640 | 52.3 | 42.7 | 5.0 |
| Asian (All) | 10,009 | 21.6 | 53.0 | 25.4 |
| Native Hawaiian or Other Pacific Islander (All) | 409 | 44.7 | 51.1 | 4.2 |
| Filipino (All) | 3,210 | 27.8 | 58.1 | 14.1 |
| Hispanic or Latino (All) | 57,049 | 49.9 | 46.0 | 4.1 |
| Black or African American (All) | 3,761 | 58.7 | 37.5 | 3.7 |
| White (All) | 26,726 | 34.3 | 54.6 | 11.1 |
| Two or more races (All) | 4,111 | 34.9 | 52.8 | 12.4 |
| Special education services | 9,352 | 72.4 | 26.3 | 1.3 |
| No special education services | 96,563 | 39.4 | 51.4 | 9.2 |
| Migrant education | 739 | 60.5 | 36.5 | 3.0 |
| Not migrant education | 105,176 | 42.2 | 49.3 | 8.5 |
| Military | 1,601 | 41.0 | 47.9 | 11.1 |
| Not military | 104,314 | 42.4 | 49.2 | 8.4 |

Table 6.D.12 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 3,444 | 54.8 | 42.3 | 3.0 |
| Not homeless | 102,471 | 41.9 | 49.4 | 8.7 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 260 | 47.7 | 45.4 | 6.9 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 380 | 55.5 | 40.8 | 3.7 |
| Asian (Primary ethnicity—not economically disadvantaged) | 6,192 | 16.3 | 51.6 | 32.2 |
| Asian (Primary ethnicity—economically disadvantaged) | 3,817 | 30.2 | 55.3 | 14.4 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 192 | 38.5 | 53.6 | 7.8 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 217 | 50.2 | 48.8 | 0.9 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 2,066 | 24.9 | 58.5 | 16.7 |
| Filipino (Primary ethnicity—economically disadvantaged) | 1,144 | 33.0 | 57.5 | 9.5 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 13,784 | 43.2 | 50.2 | 6.6 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 43,265 | 52.1 | 44.7 | 3.3 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 1,330 | 48.4 | 45.1 | 6.5 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 2,431 | 64.4 | 33.4 | 2.3 |
| White (Primary ethnicity—not economically disadvantaged) | 19,026 | 30.3 | 56.5 | 13.3 |
| White (Primary ethnicity—economically disadvantaged) | 7,700 | 44.1 | 50.1 | 5.8 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,658 | 29.2 | 54.9 | 16.0 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,453 | 45.3 | 48.9 | 5.8 |

Table 6.D.13 Percent of Students in Each Achievement Level by Demographic Variables for Grade Five—Earth and Space Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 61,607 | 42.1 | 46.6 | 11.3 |
| Male | 31,258 | 42.4 | 45.6 | 12.1 |
| Female | 30,347 | 41.9 | 47.7 | 10.5 |
| Nonbinary | 2 | N/A | N/A | N/A |
| EL | 11,181 | 75.4 | 23.6 | 1.0 |
| English only | 38,520 | 36.9 | 50.4 | 12.8 |
| RFEP | 9,510 | 30.9 | 56.9 | 12.3 |
| IFEP | 2,356 | 15.1 | 53.1 | 31.8 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 11 | 45.5 | 45.5 | 9.1 |
| English proficiency unknown | 29 | 48.3 | 37.9 | 13.8 |
| Economically disadvantaged | 33,590 | 56.5 | 39.5 | 4.0 |
| Not economically disadvantaged | 28,017 | 24.9 | 55.1 | 20.0 |
| American Indian or Alaska Native (All) | 497 | 52.7 | 41.9 | 5.4 |
| Asian (All) | 7,100 | 17.0 | 52.2 | 30.9 |
| Native Hawaiian or Other Pacific Islander (All) | 231 | 48.1 | 41.6 | 10.4 |
| Filipino (All) | 1,122 | 25.3 | 59.9 | 14.8 |
| Hispanic or Latino (All) | 29,778 | 56.0 | 39.9 | 4.1 |
| Black or African American (All) | 2,212 | 60.7 | 35.8 | 3.5 |
| White (All) | 17,091 | 29.1 | 55.5 | 15.4 |
| Two or more races (All) | 3,576 | 30.8 | 52.3 | 16.8 |
| Special education services | 6,955 | 70.6 | 25.4 | 3.9 |
| No special education services | 54,652 | 38.5 | 49.3 | 12.2 |
| Migrant education | 619 | 69.8 | 28.3 | 1.9 |
| Not migrant education | 60,988 | 41.8 | 46.8 | 11.4 |
| Military | 682 | 33.7 | 55.7 | 10.6 |
| Not military | 60,925 | 42.2 | 46.5 | 11.3 |

Table 6.D.13 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 1,442 | 64.6 | 33.2 | 2.2 |
| Not homeless | 60,165 | 41.6 | 46.9 | 11.5 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 147 | 38.8 | 52.4 | 8.8 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 350 | 58.6 | 37.4 | 4.0 |
| Asian (Primary ethnicity—not economically disadvantaged) | 5,444 | 10.0 | 52.7 | 37.3 |
| Asian (Primary ethnicity—economically disadvantaged) | 1,656 | 40.0 | 50.4 | 9.7 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 100 | 33.0 | 51.0 | 16.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 131 | 59.5 | 34.4 | 6.1 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 782 | 21.6 | 60.4 | 18.0 |
| Filipino (Primary ethnicity—economically disadvantaged) | 340 | 33.8 | 58.8 | 7.4 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 7,262 | 40.7 | 50.9 | 8.5 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 22,516 | 60.9 | 36.4 | 2.7 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 678 | 44.4 | 49.1 | 6.5 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,534 | 67.9 | 29.9 | 2.2 |
| White (Primary ethnicity—not economically disadvantaged) | 11,315 | 21.4 | 59.0 | 19.5 |
| White (Primary ethnicity—economically disadvantaged) | 5,776 | 44.1 | 48.6 | 7.3 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,289 | 21.5 | 55.8 | 22.7 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,287 | 47.6 | 46.2 | 6.3 |

Table 6.D.14 Percent of Students in Each Achievement Level by Demographic Variables for Grade Eight—Earth and Space Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 72,407 | 40.8 | 47.0 | 12.2 |
| Male | 37,042 | 41.9 | 44.7 | 13.4 |
| Female | 35,342 | 39.7 | 49.4 | 10.9 |
| Nonbinary | 23 | 34.8 | 43.5 | 21.7 |
| EL | 9,941 | 76.5 | 22.9 | 0.7 |
| English only | 41,545 | 36.7 | 49.3 | 14.0 |
| RFEP | 17,637 | 34.6 | 54.4 | 10.9 |
| IFEP | 3,248 | 17.8 | 51.3 | 30.8 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 6 | N/A | N/A | N/A |
| English proficiency unknown | 30 | 50.0 | 40.0 | 10.0 |
| Economically disadvantaged | 39,562 | 52.5 | 42.3 | 5.2 |
| Not economically disadvantaged | 32,845 | 26.8 | 52.7 | 20.5 |
| American Indian or Alaska Native (All) | 538 | 55.8 | 40.0 | 4.3 |
| Asian (All) | 7,743 | 15.7 | 52.4 | 31.9 |
| Native Hawaiian or Other Pacific Islander (All) | 259 | 46.3 | 47.1 | 6.6 |
| Filipino (All) | 1,766 | 24.0 | 58.8 | 17.2 |
| Hispanic or Latino (All) | 36,385 | 51.6 | 43.0 | 5.4 |
| Black or African American (All) | 2,603 | 58.0 | 38.0 | 4.0 |
| White (All) | 19,396 | 31.3 | 52.0 | 16.7 |
| Two or more races (All) | 3,717 | 31.4 | 50.5 | 18.1 |
| Special education services | 7,872 | 73.1 | 24.2 | 2.7 |
| No special education services | 64,535 | 36.9 | 49.8 | 13.3 |
| Migrant education | 641 | 61.5 | 36.3 | 2.2 |
| Not migrant education | 71,766 | 40.7 | 47.1 | 12.2 |
| Military | 1,028 | 32.6 | 52.9 | 14.5 |
| Not military | 71,379 | 41.0 | 46.9 | 12.1 |

Table 6.D.14 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 1,661 | 59.4 | 37.3 | 3.4 |
| Not homeless | 70,746 | 40.4 | 47.2 | 12.4 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 181 | 48.1 | 43.1 | 8.8 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 357 | 59.7 | 38.4 | 2.0 |
| Asian (Primary ethnicity—not economically disadvantaged) | 5,783 | 9.6 | 52.8 | 37.6 |
| Asian (Primary ethnicity—economically disadvantaged) | 1,960 | 33.6 | 51.2 | 15.2 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 116 | 31.0 | 60.3 | 8.6 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 143 | 58.7 | 36.4 | 4.9 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 1,210 | 20.7 | 59.1 | 20.2 |
| Filipino (Primary ethnicity—economically disadvantaged) | 556 | 31.3 | 58.1 | 10.6 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 9,002 | 40.0 | 49.8 | 10.2 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 27,383 | 55.4 | 40.8 | 3.9 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 842 | 43.9 | 48.5 | 7.6 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,761 | 64.7 | 32.9 | 2.3 |
| White (Primary ethnicity—not economically disadvantaged) | 13,275 | 25.3 | 54.2 | 20.6 |
| White (Primary ethnicity—economically disadvantaged) | 6,121 | 44.5 | 47.3 | 8.2 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,436 | 23.1 | 53.2 | 23.7 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,281 | 47.2 | 45.4 | 7.4 |

Table 6.D.15 Percent of Students in Each Achievement Level by Demographic Variables for Grade Ten—Earth and Space Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 2,233 | 33.9 | 48.9 | 17.2 |
| Male | 1,117 | 35.5 | 45.1 | 19.3 |
| Female | 1,115 | 32.3 | 52.6 | 15.1 |
| Nonbinary | 1 | N/A | N/A | N/A |
| EL | 153 | 83.7 | 15.7 | 0.7 |
| English only | 1,425 | 28.5 | 54.2 | 17.3 |
| RFEP | 512 | 39.8 | 47.9 | 12.3 |
| IFEP | 143 | 13.3 | 35.7 | 51.0 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 0 | N/A | N/A | N/A |
| English proficiency unknown | 0 | N/A | N/A | N/A |
| Economically disadvantaged | 983 | 49.9 | 44.9 | 5.2 |
| Not economically disadvantaged | 1,250 | 21.3 | 52.1 | 26.6 |
| American Indian or Alaska Native (All) | 22 | 59.1 | 36.4 | 4.5 |
| Asian (All) | 307 | 8.5 | 35.5 | 56.0 |
| Native Hawaiian or Other Pacific Islander (All) | 7 | N/A | N/A | N/A |
| Filipino (All) | 19 | 15.8 | 63.2 | 21.1 |
| Hispanic or Latino (All) | 882 | 50.7 | 45.0 | 4.3 |
| Black or African American (All) | 52 | 42.3 | 46.2 | 11.5 |
| White (All) | 821 | 26.1 | 57.9 | 16.1 |
| Two or more races (All) | 123 | 24.4 | 51.2 | 24.4 |
| Special education services | 221 | 68.3 | 26.2 | 5.4 |
| No special education services | 2,012 | 30.1 | 51.4 | 18.5 |
| Migrant education | 13 | 69.2 | 30.8 | 0.0 |
| Not migrant education | 2,220 | 33.7 | 49.0 | 17.3 |
| Military | 14 | 42.9 | 42.9 | 14.3 |
| Not military | 2,219 | 33.8 | 48.9 | 17.2 |

Table 6.D.15 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 28 | 71.4 | 28.6 | 0.0 |
| Not homeless | 2,205 | 33.4 | 49.2 | 17.4 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 9 | N/A | N/A | N/A |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 13 | 69.2 | 30.8 | 0.0 |
| Asian (Primary ethnicity—not economically disadvantaged) | 279 | 5.0 | 36.6 | 58.4 |
| Asian (Primary ethnicity—economically disadvantaged) | 28 | 42.9 | 25.0 | 32.1 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 6 | N/A | N/A | N/A |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 1 | N/A | N/A | N/A |
| Filipino (Primary ethnicity—not economically disadvantaged) | 16 | 12.5 | 62.5 | 25.0 |
| Filipino (Primary ethnicity—economically disadvantaged) | 3 | N/A | N/A | N/A |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 272 | 38.2 | 53.3 | 8.5 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 610 | 56.2 | 41.3 | 2.5 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 26 | 34.6 | 50.0 | 15.4 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 26 | 50.0 | 42.3 | 7.7 |
| White (Primary ethnicity—not economically disadvantaged) | 557 | 20.8 | 59.4 | 19.7 |
| White (Primary ethnicity—economically disadvantaged) | 264 | 37.1 | 54.5 | 8.3 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 85 | 18.8 | 49.4 | 31.8 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 38 | 36.8 | 55.3 | 7.9 |

Table 6.D.16 Percent of Students in Each Achievement Level by Demographic Variables for Grade Eleven—Earth and Space Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 43,005 | 37.3 | 49.1 | 13.6 |
| Male | 21,441 | 38.8 | 46.1 | 15.1 |
| Female | 21,546 | 35.8 | 52.0 | 12.2 |
| Nonbinary | 18 | 27.8 | 55.6 | 16.7 |
| EL | 3,290 | 79.4 | 20.3 | 0.3 |
| English only | 23,845 | 33.1 | 50.7 | 16.1 |
| RFEP | 14,020 | 36.8 | 52.7 | 10.5 |
| IFEP | 1,841 | 20.0 | 50.8 | 29.2 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 1 | N/A | N/A | N/A |
| English proficiency unknown | 8 | N/A | N/A | N/A |
| Economically disadvantaged | 23,211 | 45.5 | 47.0 | 7.4 |
| Not economically disadvantaged | 19,794 | 27.7 | 51.4 | 20.9 |
| American Indian or Alaska Native (All) | 302 | 44.0 | 48.0 | 7.9 |
| Asian (All) | 4,444 | 19.2 | 48.9 | 31.8 |
| Native Hawaiian or Other Pacific Islander (All) | 169 | 41.4 | 47.9 | 10.7 |
| Filipino (All) | 1,301 | 20.4 | 58.0 | 21.5 |
| Hispanic or Latino (All) | 21,080 | 46.2 | 47.2 | 6.5 |
| Black or African American (All) | 1,357 | 51.6 | 41.7 | 6.7 |
| White (All) | 12,590 | 29.6 | 51.9 | 18.4 |
| Two or more races (All) | 1,762 | 31.0 | 49.7 | 19.3 |
| Special education services | 3,621 | 72.5 | 24.8 | 2.7 |
| No special education services | 39,384 | 34.1 | 51.3 | 14.6 |
| Migrant education | 264 | 57.2 | 40.5 | 2.3 |
| Not migrant education | 42,741 | 37.2 | 49.1 | 13.7 |
| Military | 639 | 31.3 | 47.9 | 20.8 |
| Not military | 42,366 | 37.4 | 49.1 | 13.5 |

Table 6.D.16 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 1,323 | 51.5 | 44.1 | 4.4 |
| Not homeless | 41,682 | 36.9 | 49.2 | 13.9 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 137 | 38.7 | 51.1 | 10.2 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 165 | 48.5 | 45.5 | 6.1 |
| Asian (Primary ethnicity—not economically disadvantaged) | 2,726 | 13.7 | 46.4 | 39.9 |
| Asian (Primary ethnicity—economically disadvantaged) | 1,718 | 28.1 | 52.9 | 19.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 81 | 27.2 | 56.8 | 16.0 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 88 | 54.5 | 39.8 | 5.7 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 860 | 17.7 | 58.6 | 23.7 |
| Filipino (Primary ethnicity—economically disadvantaged) | 441 | 25.9 | 56.9 | 17.2 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 5,422 | 39.0 | 50.4 | 10.6 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 15,658 | 48.7 | 46.2 | 5.1 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 508 | 44.5 | 45.1 | 10.4 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 849 | 55.8 | 39.7 | 4.5 |
| White (Primary ethnicity—not economically disadvantaged) | 8,925 | 25.3 | 53.1 | 21.6 |
| White (Primary ethnicity—economically disadvantaged) | 3,665 | 40.1 | 49.2 | 10.7 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 1,135 | 24.0 | 52.5 | 23.5 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 627 | 43.9 | 44.5 | 11.6 |

Table 6.D.17 Percent of Students in Each Achievement Level by Demographic Variables for Grade Twelve—Earth and Space Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 60,746 | 40.3 | 48.5 | 11.1 |
| Male | 29,930 | 40.4 | 46.8 | 12.8 |
| Female | 30,794 | 40.2 | 50.2 | 9.5 |
| Nonbinary | 22 | 40.9 | 50.0 | 9.1 |
| EL | 4,936 | 80.0 | 19.5 | 0.4 |
| English only | 30,616 | 35.5 | 50.3 | 14.2 |
| RFEP | 22,061 | 39.8 | 52.2 | 8.0 |
| IFEP | 3,056 | 27.1 | 51.9 | 21.0 |
| ADEL | 68 | 80.9 | 19.1 | 0.0 |
| To be determined | 2 | N/A | N/A | N/A |
| English proficiency unknown | 7 | N/A | N/A | N/A |
| Economically disadvantaged | 36,265 | 47.6 | 46.4 | 6.0 |
| Not economically disadvantaged | 24,481 | 29.6 | 51.7 | 18.7 |
| American Indian or Alaska Native (All) | 316 | 44.6 | 48.4 | 7.0 |
| Asian (All) | 5,260 | 20.9 | 51.2 | 27.9 |
| Native Hawaiian or Other Pacific Islander (All) | 233 | 42.5 | 51.5 | 6.0 |
| Filipino (All) | 1,890 | 25.6 | 58.5 | 15.9 |
| Hispanic or Latino (All) | 35,138 | 47.7 | 46.6 | 5.7 |
| Black or African American (All) | 2,359 | 56.5 | 38.8 | 4.7 |
| White (All) | 13,323 | 29.2 | 52.4 | 18.4 |
| Two or more races (All) | 2,227 | 31.7 | 50.8 | 17.6 |
| Special education services | 5,512 | 70.6 | 26.8 | 2.6 |
| No special education services | 55,234 | 37.3 | 50.7 | 12.0 |
| Migrant education | 464 | 58.2 | 38.4 | 3.4 |
| Not migrant education | 60,282 | 40.2 | 48.6 | 11.2 |
| Military | 950 | 39.5 | 47.6 | 12.9 |
| Not military | 59,796 | 40.3 | 48.5 | 11.1 |

Table 6.D.17 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 2,095 | 51.9 | 43.9 | 4.2 |
| Not homeless | 58,651 | 39.9 | 48.7 | 11.4 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 114 | 37.7 | 55.3 | 7.0 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 202 | 48.5 | 44.6 | 6.9 |
| Asian (Primary ethnicity—not economically disadvantaged) | 3,187 | 15.2 | 50.3 | 34.5 |
| Asian (Primary ethnicity—economically disadvantaged) | 2,073 | 29.8 | 52.6 | 17.6 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 105 | 36.2 | 56.2 | 7.6 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 128 | 47.7 | 47.7 | 4.7 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 1,191 | 23.5 | 58.0 | 18.5 |
| Filipino (Primary ethnicity—economically disadvantaged) | 699 | 29.0 | 59.4 | 11.6 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 8,098 | 40.2 | 50.0 | 9.8 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 27,040 | 49.9 | 45.6 | 4.5 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 798 | 45.5 | 47.2 | 7.3 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 1,561 | 62.1 | 34.5 | 3.4 |
| White (Primary ethnicity—not economically disadvantaged) | 9,550 | 25.2 | 53.2 | 21.6 |
| White (Primary ethnicity—economically disadvantaged) | 3,773 | 39.2 | 50.4 | 10.4 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 1,438 | 26.4 | 51.1 | 22.5 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 789 | 41.2 | 50.2 | 8.6 |

Table 6.D.18 Percent of Students in Each Achievement Level by Demographic Variables for High School—Earth and Space Sciences Domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| All students | 105,984 | 39.0 | 48.7 | 12.3 |
| Male | 52,488 | 39.7 | 46.5 | 13.9 |
| Female | 53,455 | 38.3 | 51.0 | 10.7 |
| Nonbinary | 41 | 34.1 | 53.7 | 12.2 |
| EL | 8,379 | 79.9 | 19.8 | 0.4 |
| English only | 55,886 | 34.3 | 50.6 | 15.1 |
| RFEP | 36,593 | 38.7 | 52.3 | 9.0 |
| IFEP | 5,040 | 24.1 | 51.0 | 24.9 |
| ADEL | 68 | 80.9 | 19.1 | 0.0 |
| To be determined | 3 | N/A | N/A | N/A |
| English proficiency unknown | 15 | 53.3 | 46.7 | 0.0 |
| Economically disadvantaged | 60,459 | 46.8 | 46.6 | 6.6 |
| Not economically disadvantaged | 45,525 | 28.5 | 51.6 | 19.9 |
| American Indian or Alaska Native (All) | 640 | 44.8 | 47.8 | 7.3 |
| Asian (All) | 10,011 | 19.8 | 49.7 | 30.5 |
| Native Hawaiian or Other Pacific Islander (All) | 409 | 41.8 | 50.1 | 8.1 |
| Filipino (All) | 3,210 | 23.4 | 58.3 | 18.2 |
| Hispanic or Latino (All) | 57,100 | 47.2 | 46.8 | 6.0 |
| Black or African American (All) | 3,768 | 54.5 | 39.9 | 5.5 |
| White (All) | 26,734 | 29.3 | 52.3 | 18.4 |
| Two or more races (All) | 4,112 | 31.2 | 50.3 | 18.5 |
| Special education services | 9,354 | 71.3 | 26.0 | 2.7 |
| No special education services | 96,630 | 35.8 | 51.0 | 13.2 |
| Migrant education | 741 | 58.0 | 39.0 | 3.0 |
| Not migrant education | 105,243 | 38.8 | 48.8 | 12.3 |
| Military | 1,603 | 36.2 | 47.7 | 16.1 |
| Not military | 104,381 | 39.0 | 48.8 | 12.2 |

Table 6.D.18 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Below Standard | Near Standard | Above Standard |
| Homeless | 3,446 | 51.9 | 43.8 | 4.2 |
| Not homeless | 102,538 | 38.5 | 48.9 | 12.6 |
| American Indian or Alaska Native (Primary ethnicity—not economically disadvantaged) | 260 | 38.5 | 52.7 | 8.8 |
| American Indian or Alaska Native (Primary ethnicity—economically disadvantaged) | 380 | 49.2 | 44.5 | 6.3 |
| Asian (Primary ethnicity—not economically disadvantaged) | 6,192 | 14.1 | 48.0 | 38.0 |
| Asian (Primary ethnicity—economically disadvantaged) | 3,819 | 29.1 | 52.6 | 18.4 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—not economically disadvantaged) | 192 | 31.8 | 56.8 | 11.5 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—economically disadvantaged) | 217 | 50.7 | 44.2 | 5.1 |
| Filipino (Primary ethnicity—not economically disadvantaged) | 2,067 | 21.0 | 58.3 | 20.7 |
| Filipino (Primary ethnicity—economically disadvantaged) | 1,143 | 27.8 | 58.4 | 13.7 |
| Hispanic or Latino (Primary ethnicity—not economically disadvantaged) | 13,792 | 39.7 | 50.2 | 10.1 |
| Hispanic or Latino (Primary ethnicity—economically disadvantaged) | 43,308 | 49.6 | 45.8 | 4.7 |
| Black or African American (Primary ethnicity—not economically disadvantaged) | 1,332 | 44.9 | 46.5 | 8.6 |
| Black or African American (Primary ethnicity—economically disadvantaged) | 2,436 | 59.8 | 36.4 | 3.8 |
| White (Primary ethnicity—not economically disadvantaged) | 19,032 | 25.1 | 53.3 | 21.5 |
| White (Primary ethnicity—economically disadvantaged) | 7,702 | 39.5 | 50.0 | 10.5 |
| Two or more races (Primary ethnicity—not economically disadvantaged) | 2,658 | 25.1 | 51.7 | 23.2 |
| Two or more races (Primary ethnicity—economically disadvantaged) | 1,454 | 42.2 | 47.9 | 9.9 |

## 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) calibration, and response time analyses, as well as analyses to support reliability and validity evidence. A brief summary of the analyses to compare results of testing location (i.e., remote versus in person) are also included in this chapter. All results presented are for students who participated in the computer-based testing, as only six students tested on paper (two students each in grades five and eight and in high school) during the 2020–2021 administration.

### Sample Used for Analyses

Two item analyses were run for the CAST: the preliminary item analysis (PIA) and the final item analysis (FIA).

PIA identifies potentially problematic items for further evaluation and is run as soon as a sufficient volume of data is collected to obtain stable statistics. In the 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 [*6.1.1.1.2 Sampling Process for Field Te**st C**onstructed-Response Items*](#_Sampling_Process_for) for details). The PIA included data from both machine-scored and CR items as a single data set for each grade level or grade band.

FIA is conducted near the end of a typical administration. 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 the test 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.
* 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 test but did not complete it during the testing window.

For score reporting, missing responses for the machine-scorable items because of “omit” were treated as “incorrect.” Not-reached items were not included in the calculation of student scores.

Any field test items flagged during the PIA were sent to the data review committee (refer to section [*3.6 Data Review Meeting*](#_Data_Review_Meeting) for more details) for evaluation. The California Department of Education (CDE) then made final decisions on the acceptance or rejection of the items based on the data review results. Items that were rejected by the CDE were not included in the FIA and the IRT calibration process.

### Classical Item Analyses

Classical item analyses are conducted to evaluate the performance of all test items with respect to item difficulty, item discrimination, and distractor analysis. In addition, the distributions of score categories on key-based, selected-response items and rule-based, machine-scored items are also included in the classical item analyses results. Lastly, the associated flagging rules of these statistics are used to identify items that are not performing as expected.

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. Table 7.1 and table 7.2 present the item-by-item proportion-correct indices (*p*-values) and the item-total correlation indices for both dichotomous and polytomous items, respectively. The *p-*values by item type and content domain are presented in [appendix 7.A](#_Appendix_7.A:_Item); item-total correlations by item and type and content domain are presented in [appendix 7.B](#_Appendix_7.B:_Item-Total).

The omit rate of items, distractor analysis, and the distributions of score categories for the polytomous items were also included in the classical item analyses results. Lastly, the associated flagging rules of these statistics were used to identify items that were not performing well.

#### Classical Item Difficulty Indices (*p*-value)

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.

The formula for the *p*-value for a dichotomous item is presented in equation 7.1. *Refer to the* [*Alternative Text for Equation 7.1*](#_Alternative_Text_for) *for a description of this equation.*

 (7.1)

where,

*Xic* is the number of students who answered item *i* correctly, and

*Ni* is the total number of students who were presented with item *i*.

For polytomous items, the difficulty is indicated by the average item score (AIS) or *p*-value. The AIS can range from 0.00 to the maximum total possible points for an item. 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 7.2. *Refer to the* [*Alternative Text for Equation 7.2*](#_Alternative_Text_for_1) *for a description of this equation.*

 (7.2)

where,

*Xij* is the score assigned for a given polytomous item *i* and student *j*,

*Ni* is the total number of students who were presented with item *i*, and

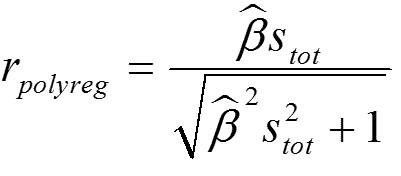
*Max (Xi)* is the maximum possible score for item *i*.

Acceptable *p*-values for both dichotomous and polytomous items for the CAST are between 0.20 and 0.95. Items with *p*-values outside this range were flagged for review.

#### Item-Total Correlation

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 operational scale 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 7.3. *Refer to the* [*Alternative Text for Equation 7.3*](#_Alternative_Text_for_2) *for a description of this equation.*

 (7.3)

where,

 is the item parameter to be estimated from the data, with the estimate denoted as , using maximum likelihood estimation; it is a regression coefficient (slope) for predicting the continuous version of an item score onto the continuous version of the total score; and

*Stot* is the standard deviation (SD) of the criterion (the students’ total score).

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 *k-*1 regressions. Beta (*β*) is the common slope for all *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 students with higher total raw scores on the overall test tend to perform better on the item than students with lower total raw scores. An item with a negative item-total correlation typically signifies a problem with the item, as that indicates that

* the higher-ability students on the overall test tend to respond incorrectly to the item if dichotomous, or are assigned a low score for the item if polytomous; or
* the lower-ability students on the overall test 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 no credit, the rubric may be inappropriate for the item.

Items with a low percentage (i.e., less than 3 percent) of students obtaining any score point were flagged for review. Such items may pose problems during the IRT calibrations. They need to be carefully reviewed and may need to be excluded from the item calibration analyses.

#### Omit Rates

An item is considered “omitted” if it was seen but not answered (i.e., it was left blank). Because the CAST required students to provide answers to all items on a page before moving on to the next page, the possibility of an omission would be very small.

#### Distractor Analyses

##### The Proportion of Students Choosing Each Distractor

For the CAST, distractor analyses were conducted on selected-response (SR) items (i.e., items that were not CRs). 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 (i.e., the item was miskeyed).

##### Distractor-Total Correlation

For SR items, the distractor-total correlation describes the relationship between selecting a distractor for a specific item and performance on the total test. 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 positive distractor-total correlations 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

In summary, items were flagged for review if the item analysis yielded any of the following results:

* **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 value less than 0.2 suggests that the item might be too difficult.

A value greater than 0.95 suggests that the item might be too easy.

* A **discrimination flag** indicated that the item does not discriminate effectively between high- and low-ability students. Items with an item-total polyserial correlation less than 0.20 are flagged.
* An **omit flag** was set if the nonresponse rates were greater than 5 percent for both dichotomous and polytomous items.
* A **distractor flag** was used for any distractors having a positive correlation with the criterion score.
* A **miskey flag** was used for SR items when more of the high-ability student group—the top 20 percent of students on the total test—chose any distractor rather than the response keyed as correct.
* The **underrepresented score point flag** was used for any item that had less than 3 percent of the students at any score level.

ETS’ Psychometric Analysis & Research staff and Assessment and Learning Technology Research & Development staff reviewed each of the flagged items at the end of the item analyses and summarized the results for the CDE.

#### 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 7.1. Across grade levels, most items had *p*-value between 0.2 and 0.8 and only a few were outside of this range. Item difficulty distributions by item type are shown in table 7.A.1 in [appendix 7.A](#_Appendix_7.A:_Item); item difficulty distributions by content domain are presented in table 7.A.2.

Table 7.1 Item Difficulty Distributions

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | 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 | 1 | 15 | 35 | 17 | 1 | 69 |
| Grade 8 | 1 | 31 | 34 | 3 | 0 | 69 |
| High school—Grade 10 | 3 | 18 | 27 | 13 | 1 | 62 |
| High school—Grade 11 | 3 | 19 | 30 | 10 | 0 | 62 |
| High school—Grade 12 | 6 | 20 | 27 | 9 | 0 | 62 |
| High school—All grades | 4 | 21 | 27 | 10 | 0 | 62 |

Overall item-total correlation distributions are presented in table 7.2. Across grade levels, the item-total correlations were 0.2 or higher. No item-total correlations were negative. Item-total correlation distributions by item type are shown in table 7.B.1 in [appendix 7.B](#_Appendix_6.D_Item); item-total correlation distributions by content domain are presented in table 7.B.2.

Table 7.2 Item-Total Correlation Distributions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Grade Level | 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 | 1 | 6 | 13 | 49 | 69 |
| Grade 8 | 0 | 2 | 3 | 7 | 12 | 45 | 69 |
| High school—Grade 10 | 0 | 2 | 1 | 4 | 11 | 44 | 62 |
| High school—Grade 11 | 0 | 3 | 0 | 7 | 13 | 39 | 62 |
| High school—Grade 12 | 0 | 3 | 0 | 8 | 11 | 40 | 62 |
| High school—All grades | 0 | 3 | 0 | 8 | 11 | 40 | 62 |

### Differential Item Functioning Analyses

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 test.

DIF analyses were conducted for CAST 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 (i.e., 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 (i.e., 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.

#### 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 to relevant student groups of interest (e.g., gender or ethnicity). Students at each total score level in the focal group (e.g., females) are compared with students at each total score level in the reference group (e.g., 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 7.4 (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 7.4*](#_Alternative_Text_for_3) *for a description of this equation.*

Equation 7.4; a link to the long description for this equation is found in the preceding paragraph. (7.4)

where,

*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 to the delta scale using equation 7.5 (Holland & Thayer, 1985). Refer to the *[Alternative Text for Equation 7.5](#_Alternative_Text_for_4)* for a description of this equation.

 (7.5)

Positive values indicate DIF in favor of the focal group (i.e., positive DIF items are differentially easier for the focal group), whereas negative values indicate DIF in favor of the reference group (i.e., 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 7.6. Refer to the [*Alternative Text for Equation 7.6*](#_Alternative_Text_for_5)for a description of this equation.

 (7.6)

where,

*X* isthe criterion score (total raw score),

*Y* isthe item score,

*M* is the number of score levels on *X*,

*Nfm* is the number of students in the focal group at score level *m*,

*Er* is the expected item score for the reference group,

*Ef* is the expected item score for the focal group, 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 (i.e., 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 (i.e., the item is differentially harder for the focal group).

#### Classification

Based on the DIF statistics 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.

The flagging criteria for dichotomous items are presented in table 7.3; the flagging criteria for polytomous items are provided in table 7.4.

Table 7.3 DIF Categories for Dichotomous Items

|  |  |
| --- | --- |
| DIF Category | Criteria |
| A (negligible) | * Absolute value of MH D-DIF is not significantly different from zero or is less than one. * 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 significantly different from one and is at least 1.5. * Positive values are classified as “C+” and negative values as “C-.” |

Table 7.4 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 |

**Note:** SMD = standardized mean difference; SD = total group standard deviation of item score

DIF analyses were conducted on each test for designated comparison groups. Groups are defined on the basis of demographic variables, including 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 7.5.

Table 7.5 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 |

Table 7.5 *(continuation)*

|  |  |  |
| --- | --- | --- |
| DIF Type | Focal Group | Reference Group |
| English fluency | English learner (EL) | English only |
| Disability | Special education services | No special education services |
| Economic status | Economically disadvantaged | Not economically disadvantaged |

#### Differential Item Functioning Analysis Results

Summarized DIF results are presented in table 7.6 through table 7.8 for grades five and eight and for high school, respectively.

No items were identified with C-level DIF in grade five. In grade eight, one item was flagged for C-level DIF, while two items were flagged for C-level DIF in high school. 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.

Test 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 7.6 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 | 68 | 0 | 1 | 0 | 0 | 0 |
| Asian–White | 65 | 2 | 2 | 0 | 0 | 0 |
| Black–White | 69 | 0 | 0 | 0 | 0 | 0 |
| Hispanic–White | 68 | 1 | 0 | 0 | 0 | 0 |
| American Indian or Alaska Native–White | 69 | 0 | 0 | 0 | 0 | 0 |
| EL–English only | 69 | 0 | 0 | 0 | 0 | 0 |
| Special education services–No special education services | 69 | 0 | 0 | 0 | 0 | 0 |
| Economically disadvantaged–Not economically disadvantaged | 69 | 0 | 0 | 0 | 0 | 0 |

Table 7.7 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 | 67 | 1 | 0 | 1 | 0 | 0 |
| Asian–White | 69 | 0 | 0 | 0 | 0 | 0 |
| Black–White | 69 | 0 | 0 | 0 | 0 | 0 |
| Hispanic–White | 67 | 0 | 2 | 0 | 0 | 0 |
| American Indian or Alaska Native–White | 67 | 0 | 2 | 0 | 0 | 0 |
| EL–English only | 68 | 1 | 0 | 0 | 0 | 0 |
| Special education services–No special education services | 69 | 0 | 0 | 0 | 0 | 0 |
| Economically disadvantaged–Not economically disadvantaged | 69 | 0 | 0 | 0 | 0 | 0 |

Table 7.8 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 | 60 | 0 | 1 | 0 | 1 | 0 |
| Asian–White | 61 | 1 | 0 | 0 | 0 | 0 |
| Black–White | 62 | 0 | 0 | 0 | 0 | 0 |
| Hispanic–White | 61 | 0 | 1 | 0 | 0 | 0 |
| American Indian or Alaska Native–White | 62 | 0 | 0 | 0 | 0 | 0 |
| EL–English only | 56 | 4 | 1 | 1 | 0 | 0 |
| Special education services–No special education services | 62 | 0 | 0 | 0 | 0 | 0 |
| Economically disadvantaged–Not economically disadvantaged | 62 | 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 test taker’s true ability. 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 are used with CAST data to calibrate and link field test items onto the base IRT scale established during the 2018–2019 administration. Only items that were not rejected by both the data review committees and the CDE were included in the calibration process.

The purpose of the IRT calibration and equating for the CAST is to provide item parameter estimates that are on the same scale for score reporting and future form assembly. For details on scale scores and achievement levels reported for the CAST, refer to [*Chapter 6: Sco**ring and Reporting*](#_Scoring_and_Reporting).

#### Models

On the basis of the results from the test dimensionality study conducted using data from the 2018–2019 administration (refer to *chapter 12* of the *California Science Test 2018–2019 Technical Report* [CDE, 2020]), a unidimensional model was used to calibrate the CAST items. The two-parameter logistic item response theory (2PL‑IRT) model was used to calibrate the dichotomous items (i.e., items worth 1 point) and the generalized partial credit model (GPCM) (Muraki, 1992) was used to calibrate the polytomous items (i.e., items worth more than 1 point). The 2PL-IRT model is a special case of the GPCM when the maximum number of score points for the item is 1. FlexMIRT® (Cai, 2017), a multilevel and multiple-group IRT software package (version 3.51), is used for the calibration.

The mathematical form of the GPCM (Muraki, 1992) is presented in equation 7.7. *Refer to the* [*Alternative Text for Equation 7.7*](#_Alternative_Text_for_6) *for a description of this equation.*

P sub I h of theta sub j equals the numerator exp open parenthesis the sum from v equals 1 to h of D times a sub i of the quantity open parenthesis theta sub j minus b sub I plus d sub iv close parenthesis close parenthesis and denominator 1 plus the sum from c equals 1 to n sub I exp open parenthesis the sum from v equals 1 to c D times a sub i of the quantity open parenthesis theta sub j minus b sub I plus d sub iv close parenthesis close parenthesis, if score h equals 1, 2, …., n sub i.

P sub I h of theta sub j equals 1 divided by denominator 1 plus the sum from c equals 1 to n sub I exp open parenthesis the sum from v equals 1 to c D times a sub i of the quantity open parenthesis theta sub j minus b sub I plus d sub iv close parenthesis close parenthesis, if score h equals 0. (7.7)

where,

*Pih(θj)* is the probability of student with proficiency *θj* obtaining score *h* on item *i,*

*ni* is the maximum number of score points for item *i*,

*ai* is the discrimination parameter for item *i*,

*bi* is the location parameter for item *i*,

*div* is the category parameter for item *i* on score *v*, and

*D* is a scaling constant of 1.7 that makes the logistic model approximate the normal ogive model.

When *ni* = 1, equation 7.7 becomes an expression of the two-parameter logistic model for dichotomous items.

#### Data Preparation

Following a typical administration, operational item responses and field test item responses are combined into a sparse matrix for concurrent calibration. Field test items that have been rejected by the CDE’s final decision, informed by the data review committee, are excluded from the calibration.

#### Calibration

Operational item responses and field test item responses were combined into a sparse matrix for concurrent calibration. For grade five and grade eight, data was available only for the discrete items. Data from the CR field test items for these grades had not been analyzed in time to be included in this technical report. The sample used in the item calibration included all students who took the CAST during the 2020–2021 administration.

Similar to the classical item analyses, “omit” items were treated as incorrect. The “not‑administered” items and the field test CR items that were administered but not scored were treated as not presented.

The calibration for the high school assessment was conducted using multigroup analyses, where the mean and variance of the ability estimates were set to 0 and 1 for grade eleven to serve as the reference grade level, and freely estimated for grades ten and twelve. This calibration method was used for the high school data because there might be some difference in the ability distributions for grades ten, eleven, and twelve. These differences may be due to grade ten having relatively few students taking the CAST compared to students in grades eleven and twelve (refer to table 5.1), and it is not a typical sample—usually those who are good at the subject will take the CAST early. Additionally, grade twelve students are in their senior year and motivation could be an issue. The difference in the theta distributions for the three high school grades could then be evaluated to observe any differences between the grade levels.

The item parameter estimates—the item discrimination, location, and categories parameters—were set to be equal across the three grade levels. The calibration for grades five and eight were conducted using single-group analyses.

The FlexMIRT output was then evaluated to examine whether every execution of FlexMIRT converged. The item parameter estimates were examined for reasonableness. Items with unreasonably large parameter values or standard errors were noted. Such items would not be eligible for use in future forms based on the statistical specifications for form assembly.

#### Equating

The CAST base scale was built using the 2018–2019 data. The operational items in each subsequent administration are used as anchor items to place field test item parameter estimates onto the base scale.

The Stocking-Lord (1983) method was applied for the equating. The software STUIRT (Kim & Kolen, 2004) was employed to find the Stocking-Lord equating constants for the common items. The calibrated item parameters were then transformed onto the base form scale using those equating constants (refer to table 6.8).

#### Parameter Estimates

The overall summary of the IRT *a*-parameter estimates is shown in table 7.9. 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 are also presented.

The range of *a*-parameter estimates was between 0.09 and 1.56 across grade levels. The means of *a*-parameter estimates were 0.67, 0.61, and 0.66 for grades five and eight, and for high school, respectively. In addition, the summaries of the IRT *a*‑parameter estimates for each grade-level assessment are presented in [appendix 7.C](#_Appendix_7.C:_Item), in table 7.C.1 through table 7.C.3 by item type; and table 7.C.4 through table 7.C.6 by content domain for grades five and eight and high school, respectively.

Table 7.9 Item Discrimination Parameter Distribution by Grade Level and the Grade Band

|  |  |  |  |
| --- | --- | --- | --- |
| IRT-a Range | Grade 5 | Grade 8 | High School |
| a < 0 | 0 | 0 | 0 |
| 0 ≤ a < 0.2 | 1 | 2 | 3 |
| 0.2 ≤ a < 0.4 | 11 | 12 | 7 |
| 0.4 ≤ a < 0.6 | 15 | 21 | 20 |
| 0.6 ≤ a < 0.8 | 19 | 19 | 16 |
| 0.8 ≤ a < 1.0 | 15 | 14 | 7 |
| 1.0 ≤ a < 1.2 | 6 | 1 | 7 |
| 1.2 ≤ a < 1.4 | 2 | 0 | 1 |
| 1.4 ≤ a < 1.6 | 0 | 0 | 1 |
| 1.6 ≤ a < 1.8 | 0 | 0 | 0 |
| 1.8 ≤ a < 2.0 | 0 | 0 | 0 |
| a ≥ 2.0 | 0 | 0 | 0 |
| Minimum | 0.18 | 0.09 | 0.14 |
| Maximum | 1.24 | 1.07 | 1.56 |
| Mean | 0.67 | 0.61 | 0.66 |
| SD | 0.25 | 0.22 | 0.30 |
| **Number of Items** | **69** | **69** | **62** |

Similar information for the IRT *b*-parameter estimates is shown in table 7.10 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 grade band. The means of *b*-parameter estimates were -0.07, 0.46, and 0.56 for grades five and eight and for high school, 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 7.D](#_Appendix_7.D:_Item) in table 7.D.1 through table 7.D.3 and by content domain in table 7.D.4 through table 7.D.6 for grades five and eight, and for high school, respectively.

Table 7.10 Item Difficulty Parameter Distribution by Grade Level and the 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 | 4 | 0 | 0 |
| −1.5 ≤ b < −1.0 | 5 | 1 | 2 |
| −1.0 ≤ b < −0.5 | 11 | 5 | 7 |
| −0.5 ≤ b < 0 | 20 | 13 | 9 |
| 0 ≤ b < 0.5 | 13 | 20 | 15 |
| 0.5 ≤ b < 1.0 | 8 | 12 | 12 |
| 1.0 ≤ b < 1.5 | 6 | 10 | 7 |
| 1.5 ≤ b < 2.0 | 2 | 7 | 4 |
| 2.0 ≤ b < 2.5 | 0 | 1 | 3 |
| 2.5 ≤ b < 3.0 | 0 | 0 | 2 |
| 3.0 ≤ b < 3.5 | 0 | 0 | 0 |
| b ≥ 3.5 | 0 | 0 | 1 |
| Min | -1.84 | -1.15 | -1.12 |
| Max | 1.82 | 2.17 | 3.84 |
| Mean | -0.07 | 0.46 | 0.56 |
| SD | 0.83 | 0.74 | 1.00 |
| **Number of Items** | **69** | **69** | **62** |

### Testing Time Analyses

The 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, two PTs in Segment B, and either one PT or one block of discrete items in Segment C. The CAST is an untimed assessment. [[12]](#footnote-13)

The estimated time for students to complete the test was 60 minutes for Segment A, 40 minutes for Segment B, and 20 minutes for Segment C. The time it took students to complete a test was recorded and analyzed.

Testing time analyses were based on students who logged on to the test 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, testing 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 was typically longer than that of a PT block, the testing time for the total test in table 7.11 was broken down for students who received a PT in Segment C (i.e., two Segment A blocks + two PTs + one field test PT) and those who received a discrete block in Segment C (i.e., two Segment A blocks + two PTs + one field test discrete block). The unit of testing time is minutes.

The medians (50th percentile) are used to interpret the results because medians are less impacted by the extreme values and are, therefore, more meaningful. The median of total testing time for students who received a discrete Segment C block versus students who received a PT Segment C block was 94.5 minutes versus 89.7 minutes in grade five, 96.3 minutes versus 91.1 minutes in grade eight, and 69.4 minutes versus 64.4 minutes for the high school grade band. The corresponding total testing time for students in grades ten, eleven, and twelve was 80.8 minutes versus 73.7 minutes, 73.1 minutes versus 67.6 minutes, and 66.6 minutes versus 61.9 minutes, respectively.

The total testing time was longer for students receiving a form with a discrete Segment C block than students receiving a form with a PT Segment C block. The total testing time also decreased as the grade level increased.

Note that the criterion for students to be included in table 7.11 is that they had no “Not Seen” items.

Summaries of the times that students spent for the total test are given in table 7.11.

Table 7.11 Testing Time (in Minutes) for the Total Test

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Segment | N | Mean | SD | Min | Max | 1st Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 99th Percentile |
| Grade 5 | 2A Blocks + 2 PTs + 1 Field Test Discrete Block | 30,525 | 108.7 | 65.7 | 5.2 | 1264.8 | 20.4 | 49.8 | 69.1 | 94.5 | 130.7 | 180.2 | 350.7 |
| Grade 5 | 2A Blocks + 2 PTs + 1 Field Test PT | 30,354 | 103.7 | 63.7 | 4.2 | 1226.8 | 19.5 | 47.5 | 65.1 | 89.7 | 124.5 | 172.9 | 342.7 |
| Grade 8 | 2A Blocks + 2 PTs + 1 Field Test Discrete Block | 34,672 | 111.0 | 71.4 | 4.4 | 2363.3 | 15.5 | 46.5 | 69.0 | 96.3 | 134.0 | 187.3 | 373.3 |
| Grade 8 | 2A Blocks + 2 PTs + 1 Field Test PT | 34,913 | 105.4 | 67.1 | 4.5 | 1011.9 | 15.0 | 44.5 | 65.5 | 91.1 | 127.0 | 178.9 | 353.4 |
| High school—Grade 10 | 2A Blocks + 2 PTs + 1 Field Test Discrete Block | 1,096 | 88.3 | 49.2 | 6.4 | 468.8 | 9.7 | 39.2 | 60.0 | 80.8 | 104.8 | 138.0 | 274.3 |
| High school—Grade 10 | 2A Blocks + 2 PTs + 1 Field Test PT | 1,080 | 85.7 | 58.5 | 5.5 | 541.8 | 13.5 | 37.6 | 54.3 | 73.7 | 98.9 | 135.4 | 365.7 |
| High school—Grade 11 | 2A Blocks + 2 PTs + 1 Field Test Discrete Block | 20,946 | 80.8 | 49.7 | 3.0 | 739.5 | 8.9 | 30.5 | 50.1 | 73.1 | 100.7 | 134.8 | 254.9 |
| High school—Grade 11 | 2A Blocks + 2 PTs + 1 Field Test PT | 20,902 | 75.3 | 46.2 | 2.9 | 704.4 | 8.1 | 29.1 | 46.8 | 67.6 | 93.8 | 125.5 | 241.0 |

Table 7.11 *(continuation)*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Segment | N | Mean | SD | Min | Max | 1st Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 99th Percentile |
| High school—Grade 12 | 2A Blocks + 2 PTs + 1 Field Test Discrete Block | 29,375 | 72.9 | 42.7 | 3.4 | 962.8 | 9.1 | 28.6 | 46.1 | 66.6 | 90.7 | 120.8 | 220.1 |
| High school—Grade 12 | 2A Blocks + 2 PTs + 1 Field Test PT | 29,814 | 68.1 | 40.2 | 2.3 | 830.8 | 8.5 | 26.7 | 42.9 | 61.9 | 84.8 | 113.2 | 203.9 |
| High school—All grades | 2A Blocks + 2 PTs + 1 Field Test Discrete Block | 51,417 | 76.5 | 46.0 | 3.0 | 962.8 | 9.0 | 29.5 | 47.9 | 69.4 | 95.2 | 126.9 | 234.6 |
| High school—All grades | 2A Blocks + 2 PTs + 1 Field Test PT | 51,796 | 71.4 | 43.3 | 2.3 | 830.8 | 8.4 | 27.8 | 44.6 | 64.4 | 88.8 | 118.9 | 224.2 |

Table 7.E.1 in [appendix 7.E](#_Appendix_7.E:_Response) shows the testing time by segment. The testing time for a discrete block was longer than that for a PT block for all the percentiles considered.

Table 7.E.2 shows the testing time for four item types: multiple choice or single selection, CR, technology enhanced, and composite. Because testing time was recorded at the page level, items that were on a page with multiple items were excluded from the analysis for table 7.E.2. The testing time for a CR item was the longest for each percentile.

### Reliability Analyses

Two types of reliabilities are reported in this chapter: the reliability of the test scores and the reliability of the CR scoring.

#### Test Score Reliability

Reliability is the extent to which differences in test scores reflect true differences in the knowledge, ability, or skill being tested, rather than fluctuations due to measurement error. Thus, reliability is the consistency of the scores across conditions that do not differ systematically and only contain random measurement errors. In statistical terms, the variance in the distributions of test scores—essentially, the differences among individuals—is due partly to real differences in the knowledge, skill, or ability being tested (true variance) and due partly to measurement errors inherent in the measurement process (error variance). The reliability coefficient is an estimate of the proportion of the total variance that is true variance.

##### Marginal Reliability

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 and is sometimes operationalized as the correlation between scores on two replications of the same testing procedure.

Reliability coefficients 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. Coefficient alpha (Cronbach, 1951) is among the most common of these methodologies; however, these reliability indices are not directly applicable because the CAST has multiple forms.

Instead, an IRT-based approach called marginal reliability (Green, Bock, Humphreys, Linn, & Reckase, 1984) can be used to estimate the reliability of CAST scores. The estimates of reliability coefficients reported here are for IRT model–based ability estimates.

This reliability coefficient for theta estimates,Rho sub theta theta prime, is defined, based on a single test administration, as shown in equation 7.8. *Refer to the* [*Alternative Text for Equation 7.8*](#_Alternative_Text_for_14) *for a description of this equation.*

 (7.8)

where,

S squared sub theta is the measure of variance in ability estimates,

*θ* is the ability estimate, and

M sub SE squared sub theta is an average of the variance of the ability estimates.

The standard error of measurement (SEM) of the test on the theta scale is defined as presented in equation 7.9 *(Refer to the* [*Alternative Text for Equation 7.9*](#_Alternative_Text_for_15) *for a description of this equation.)*

 (7.9)

and the SEM of the test on the scale score metric is defined as presented in equation 7.10 *(Refer to the* [*Alternative Text for Equation 7.10*](#_Alternative_Text_for_16) *for a description of this equation.)*

 (7.10)

where,

M sub SE squared sub theta and  are the mean estimation variance of theta and scale score, respectively.

Table 7.12 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 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 0.90 to 0.92 across grade levels, indicating high levels of reliability.

Table 7.12 Summary Statistics for Scale Scores and Theta Scores, Reliability, and SEMs

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Number of Students | Reliability | Scale Score Mean | Scale Score SD | Scale Score SEM | Theta Score Mean | Theta Score SD | Theta Score SEM |
| Grade 5 | 62,129 | 0.92 | 199 | 22.24 | 5.75 | -0.08 | 1.06 | 0.31 |
| Grade 8 | 74,823 | 0.91 | 399 | 22.66 | 6.10 | -0.05 | 1.07 | 0.33 |
| Grade 10 | 2,258 | 0.92 | 605 | 22.35 | 6.08 | 0.24 | 1.10 | 0.32 |
| Grade 11 | 43,694 | 0.91 | 602 | 22.09 | 6.18 | 0.08 | 1.05 | 0.32 |
| Grade 12 | 61,607 | 0.90 | 600 | 21.34 | 6.21 | -0.01 | 1.01 | 0.32 |
| High school | 107,559 | 0.91 | 601 | 21.70 | 6.20 | 0.03 | 1.03 | 0.32 |

##### Student Group Reliabilities and Standard Errors of Measurement

The reliabilities of the total test scores were also examined for various student groups within the student population. These student groups include demographic groups, as well as subgroups of students who took both the CAST and the English Language Proficiency Assessments for California (ELPAC).

###### Reliabilities by Demographic Groups

The student groups included in these analyses were defined by gender, economic status, special education services status, accommodations for students with special education services, English language fluency, primary ethnicity, migrant status, parent military status, homeless status, and a crosstab of primary ethnicity and economic status.

Table 7.F.1 through table 7.F.6 in [appendix 7.F](#_Appendix_7.F:_Reliability) provide reliabilities, theta-based SEMs, and theta score variances for the total test scores for each student group for each grade level and for high school overall.

Note that reliabilities are reported only for samples that comprise 11 or more students. Also, in some cases, score reliabilities are not estimable and are presented in the tables as “N/A.” The reliability estimates for some of the student groups can be negative because of small variation in scale scores and large conditional standard errors of measurement for extreme score values. These negative reliabilities and their associated SEMs are presented in tables as “N/A.”

###### Reliabilities by ELPAC Performance Levels

A subset of students who took the CAST also took the Summative ELPAC, which is the required state test for English language proficiency that must be given to students whose primary language is a language other than English and are classified as English learners. The Summative ELPAC results show the overall English performance level attained by students, and the performance levels are reported as the following:

* Level 1: Minimally Developed
* 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 the 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 English language proficiency. These results are provided in table 7.F.7 of [appendix 7.F](#_Appendix_7.F:_Reliability).

##### 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.

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).

Decision accuracy describes the extent to which students are classified in the same way as they would be on the basis of the average of all possible forms of a test. Decision accuracy answers the question of how closely 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 different form of an assessment. Decision consistency answers the question of what is the agreement between the classifications based on two nonoverlapping, equally difficult forms of the test.

Decision consistency values are always lower than the corresponding decision accuracy values because in decision consistency, both of the classifications of the student are based on scores that depend on which form of the test the student took. In decision accuracy, only one of the classifications is based on a score that can vary in this way.

In each case, the proportion of classifications with exact agreement is the sum of the entries in the diagonal of the contingency table representing the bivariate distribution.

Decision accuracy at a particular threshold is estimated by partitioning the estimated bivariate distribution of true scores and observed scores (refer to table 7.13) into a two‑by‑two table, using the same threshold score on both variables. The decision accuracy statistic is the sum of the proportions in the cells representing consistent classifications—above the cut on both variables or below the cut on both variables. Decision consistency is estimated in the same way, by partitioning the estimated bivariate distribution of observed scores on two forms of the test (refer to table 7.14).

Table 7.13 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 | Correct classification | Incorrect classification |
| Reaches an achievement level | Incorrect classification | Correct classification |

Table 7.14 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 |

For a test 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 consistent classifications based on the two score distributions. The results of decision accuracy and consistency analysis for the CAST are presented, by grade levels and high school grade band, in table 7.G.1 through table 7.G.12 in [appendix 7.G](#_Appendix_7.G:_Analysis). 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 consistent classifications.

The percentages of students who were classified accurately ranged from 0.81 to 0.83 across all achievement levels and from 0.92 to 0.93 for students who met or exceeded the standards. For decision consistency, the percentages of students classified correctly ranged from 0.73 to 0.75 for all achievement levels and 0.89 to 0.90 for students who met or exceeded the standards.

#### Constructed-Response Scoring Reliability

Reliability of the CR scoring is the extent to which two different raters give consistent scores on the same response. The interrater reliability analyses typically include the percent of exact and adjacent agreement between the two raters, the kappa coefficient, and the quadratic-weighted kappa (QWK) coefficient.

##### Interrater Agreement

###### Percentage Agreement

Percentage agreement between two raters includes the percentage of exact score agreement, the percentage of adjacent score agreement, and the percentage of exact plus adjacent score agreement. Adjacent score agreement means agreement between scores that differ by just one point. 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. 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 7.15. 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.

Table 7.15 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 |

The kappa statistic is defined as presented in equation 7.11 *(Refer to the* [*Alternative Text for Equation 7.11*](#_Alternative_Text_for_17) *for a description of this equation.),*

 (7.11)

equation 7.12 *(Refer to the* [*Alternative Text for Equation 7.12*](#_Alternative_Text_for_18) *for a description of this equation.),*

 (7.12)

and equation 7.13 *(Refer to the* [*Alternative Text for Equation 7.13*](#_Alternative_Text_for_19) *for a description of this equation.).*

 (7.13)

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 also used because kappa does not take into account the degree of disagreement between raters. QWK 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. The weighted kappa coefficient is defined as presented in equation 7.14. *Refer to the* [*Alternative Text for Equation 7.14*](#_Alternative_Text_for_20) *for a description of this equation.*

 (7.14)

For QWK, the weights are as presented in equation 7.15. *Refer to the* [*Alternative Text for Equation 7.15*](#_Alternative_Text_for_21) *for a description of this equation.*

 (7.15)

##### Summary of Scoring Reliabilities

The interrater reliabilities for operational CR items are shown in table 6.3 through table 6.6; table 6.3 through table 6.5 show artificial intelligence (AI)–scored and human-scored CR items, and table 6.6 shows CR items receiving two human scores. The QWK ranged from 0.58 to 0.96 for AI-scored items and 0.62 to 0.78 for human-scored items, which indicated a moderate to high level of agreement between two raters. Detailed information on interrater reliability results can be found in subsection [*6.1.1.2.9 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). It constitutes the central notion underlying the development, administration, and scoring of a test and the uses and interpretations of test scores.

Validation is the process of accumulating evidence to support each proposed score interpretation or use. 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, test scaling and equating, scoring, reporting, and score usage.

In this section, the evidence gathered is presented to support the intended uses and interpretations of scores for the CAST. This section is organized primarily around the principles prescribed by AERA, APA, and NCME’s *Standards for Educational and Psychological Testing* (2014). These *Standards* require a clear definition of the purpose of the test, 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 response processes
3. Evidence based on internal structure
4. Evidence based on relations to other variables
5. Evidence based on consequences of testing

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

#### Evidence in the Design of the CAST

##### Purpose

The CAST is designed to measure performance on the California Next Generation Science Standards (CA NGSS). The goal of the CAST is to measure what students can do in science. The CAST covers information across the three science domains of Life Sciences, Physical Sciences, and Earth and Space Sciences.

##### Constructs to Be Measured

The 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.

Test blueprints define the procedures used to measure the domains and standards. These blueprints are provided in table 4.1 through table 4.4. They also provide an operational definition of the construct to which each set of standards refers. That is, they define, 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 blueprints control 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 the 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, based on 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 [*6.2.3 Achievement Levels*](#_Achievement_Levels). A detailed description of the uses and applications of the CAST scores is presented in [*Chapter 6: Scoring and Reporting*](#_Scoring_and_Reporting). Additional information can be found in the *CAASPP Post-Test Guide* (CDE, 2021).

The 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 the CAST, unless they are eligible to participate in the alternate assessments. Students enrolled in high school are required to take the CAST once while in high school (i.e., grade ten, eleven, or twelve [as long as the student is not repeating grade twelve]), unless they are eligible to participate in the alternate assessments.

#### Evidence Based on Test 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 tests that evaluate the interactions between curriculum frameworks, testing, and instruction (Rothman, Slattery, Vranek, & Resnick, 2002; Bhola, Impara & Buckendahl, 2003; Martone & Sireci, 2009).

##### Description of California Next Generation Science Standards

As noted in section [*1.1 Background*](#_Background), the CAST is aligned with the CA NGSS. There are three main domains at each grade level: Life Sciences, Physical Sciences, and Earth and Space 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, 2019b).

##### 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 Specific**ations*](#_Specifications).

##### Assessment Blueprints

The CAST blueprints provided in table 4.1 through table 4.4 describe the content of the science assessments for all grade levels tested and how that content is assessed. The CAST blueprints reflect the depth and breadth of the PEs of the CA NGSS. The test blueprints have information about the number of items and depth of knowledge for items associated with each assessment target. Each test is described by a single blueprint for each segment of the test. For details about CAST blueprints, please refer to subsection [*4.2.1 Test Blu**eprints*](#_Test_Blueprints).

##### Alignment Study

A strong alignment between the CAST and the CA NGSS is fundamental to the meaningful measurement of student achievement and instructional effectiveness. Alignment results demonstrate that the 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, 2019b).

##### Form Assembly Process

The content standards and blueprints are the basis for choosing items for each assessment. Additionally, item difficulty and item-total score correlations are provided to evaluate the statistical characteristics of test forms. Assembly of all CAST forms meets all the content and statistical specifications. Refer to [*Chapter 4: Test Assembly*](#_Test_Assembly) for additional information.

#### Evidence Based on 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 generally includes documentation of activities such as:

* interviews with students concerning their responses to test items (i.e., think alouds);
* systematic observations of test response behavior;
* evaluation of the criteria used by judges when scoring CRs, analysis of student item‑response‑time data, and features scored by automated algorithms; and
* evaluation of the reasoning processes students employ when solving test items (Embretson, 1983; Messick, 1989; Mislevy, 2009).

##### Analysis of Testing Time

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 a test is recorded and analyzed to build a profile describing what a typical testing event looks like for each 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 the CAST is untimed. [[13]](#footnote-14)

The descriptive statistics—e.g., the number of students, mean, SD, minimum and maximum, and percentiles—of the following time variables are computed for each grade level:

* Time required to complete each segment
* Time required to complete each item type
* Time required to complete the total test

Some cases of extremely long testing times may be attributed to students with special needs taking longer to complete the tests, or the test not being closed down properly. Therefore, mean testing times may be misleading. The medians (50th percentile) are more meaningful in the interpretation of the time comparisons because medians are less impacted by extreme values than means.

Table 7.11 provides the summary of times that students spent for the total test. Table 7.E.1 and table 7.E.2 in [appendix 7.E](#_Appendix_7.E:_Response) provide the summary of testing time by segment and by item type, respectively. These tables include various percentiles of testing times (in minutes). The median testing times were 95, 96, and 69 minutes for grades five and eight and high school, respectively. Given that the estimated testing time was two hours, approximately 50 percent of the students in grades five and eight, and 75 percent of students in high school, were able to complete the test within the estimated testing time. Detailed information on students’ testing time can be found in section [*7.5 Testing Time Analyses*](#_Testing_Time_Analyses).

##### Student Survey

The student survey questions were administered at the end of the test 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:

1. Did you learn about the topics on the test in your science classes?
2. Were any questions on the test different from the types of questions you see in science classes?
3. How hard were questions on this test compared to questions you see in science classes?
4. Do you think you will be enrolling in any more science classes in high school?

Questions one through three were for all students and question four was applicable only to high school students in grades ten and eleven.

The majority of students responded that, compared with what they were taught in their science classes,

* all or most of the topics on the test were taught,
* some items on the test were different from what they were taught, and
* questions on the test were comparable in difficulty with questions they saw in class.

The student survey results show that, in general, the CAST reflects what students were taught in the classroom. Detailed information on the student survey can be found in [*Chapter 9: Student Survey*](#_Student_Survey).

Survey questions were provided in braille for students who used the braille accommodation.

#### Evidence Based on 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 the CAST based on 2018–2019 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 test information functions.

##### Dimensionality

The CAST assesses PEs as they appear in the CA NGSS, and the PEs represent a complete integration of the three dimensions, not three dimensions that coincide together. A dimensionality study was conducted during the 2018–2019 administration to determine the factorial structure of the assessments. Results suggested the test 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 the CAST, DIF analyses are conducted to assess differences in the item performance of groups of students who differ in their demographic characteristics. For the 2020–2021 CAST administration, one item in grade eight and two items in high school were identified as having significant levels of DIF. The details on how DIF analyses are performed can be found in section [*7.3 Differential Item Functioning Analyses*](#_Differential_Item_Functioning_1).

##### Overall Reliability Estimates

The results of marginal reliability analyses on the total theta scores for the CAST are presented in table 7.12. The results indicate that the reliability estimates for the CAST total scores are high, ranging from 0.90 to 0.92 across all grade levels.

##### Student Group Reliability Estimates

The reliabilities are also examined for various student groups that differ in their demographic characteristics within the student population. The characteristics considered are gender, ethnicity, economic status, special education services status, migrant status, English language fluency, parent/guardian military status, homeless status, and ethnicity by economic status (refer to table 5.2 for the demographic student groups reported). Reliability estimates and SEM information for the total test theta scores are reported for each student group in table 7.F.1 through table 7.F.6 in [appendix 7.F](#_Appendix_7.F:_Reliability).

##### Reliability of Performance Classifications

The methodology used for estimating the reliability of classification decisions is described with the decision classification analyses in subsection [*7.6.1.3 Decision Class**ification Analyses*](#_Decision_Classification_Analyses). The results of these analyses are presented in a corresponding appendix and briefly summarized in the referenced subsection.

##### Interrater Reliability

Cohen’s Kappa statistics provide evidence of the degree to which a student’s score is consistent from one rater to another. Research has shown values of kappa between 0.41 and 0.60 exhibit moderate levels of agreement between the two ratings (Landis & Koch, 1977; Flack, Afifi, Lachenbruch, & Schouten, 1988) and that values of QWK greater than 0.70 indicate excellent agreement (Williamson, Xi, & Breyer, 2012).

The results in table 6.3 through table 6.6 in [*Chapter 6: Scoring and Reporting*](#_Scoring_and_Reporting) show the QWK ranges from 0.58 to 0.96 for AI-scored items and 0.62 to 0.78 for human-scored items, which indicate moderate to high levels of agreement between two raters.

#### Evidence Based on Relations to Smarter Balanced Test Scores

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 (e.g., test scores on other tests) to evaluate the degree to which different tests actually measure different skills and the utility of test scores for predicting specific criteria (e.g., 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 7.H.1 through table 7.H.6 in [appendix 7.H](#_Appendix_7.H:_Correlations) show these correlations for both ELA and mathematics test scores by demographic student groups. Table 7.H.1 and table 7.H.2 show data for grade five, table 7.H.3 and table 7.H.4 show data for grade eight, and table 7.H.5 and table 7.H.6 show data for grade eleven. For the total student group, higher correlations were observed in grades five and eight (0.79 or greater for both ELA and mathematics) and slightly lower correlations were observed in grade eleven (0.78 for both ELA and 0.74 for mathematics).

#### Evidence Based on 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 tests, the *Standards* stress the importance of evaluating test consequences. For example, they state the following:

“When educational testing programs are mandated by school, district, state, or other authorities, 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 use 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.

### Score Comparability—Remote Versus In-Person Testing Analyses

At the start of the 2020–2021 school year, LEAs offered varying instructional options, with a substantial percentage offering only distance learning options. The CDE allowed LEAs flexibility to use multiple test administration options so LEAs could best meet the needs of students in response to the local context and to ensure the safety and health of students and LEA staff. As a result, both in-person and remote testing modes were used for the 2020–2021 CAST administration.

To evaluate whether those assessment flexibilities used in the administration impact test score interpretation and to examine whether there were issues in student test experience related to those assessment flexibilities that could affect score validity, ETS conducted a test location comparability study. This study investigated and analyzed a variety of factors that could have an impact on student testing performance and experiences, such as student test participation, student performance at the item level and the test level, student response time for the entire test, test reliability, and student testing issues related to the validity of score interpretation. The evaluation included comparisons of the performance of students testing in person and remotely at the item level and the total test level, as well as an examination of response time and test reliability.

This section provides a brief summary of the analysis sample and the results from the analyses conducted.

#### Analysis Sample

Approximately 60 percent of students tested remotely and 40 percent tested in person across the grade levels. More grade five students tested in person compared to the other grade levels.

To facilitate direct comparisons between students who tested in person and students who tested remotely, a weighting approach was used to match students within these groups to the 2018–2019 CAST population. In addition, for many students, the mode of instruction was intertwined with their testing location, which made it difficult to determine whether any differences in the students’ performance were a result of the test location difference or the instructional location difference. To provide an unconfounded comparison of test performance between students testing remotely and in person, analyses were conducted using two sets of student data:

1. All students testing in person or remotely regardless of their instructional mode during the 2020–2021 school year
2. A subset of students who received only remote instruction during the 2020–2021 school year

To eliminate the potential impact of instructional mode, the results presented will focus on the subset of students who consistently received remote instruction. The number of students by instructional mode and assessment location is presented in table 7.16.

Table 7.16 Number of Students by Instructional Mode and Assessment Location

|  |  |  |  |
| --- | --- | --- | --- |
| Grade Level | Instructional Mode | In-Person Assessment | Remote Assessment |
| Grade 5 | Only remote | 3,407 | 10,102 |
| Grade 5 | Only in person | 556 | 4 |
| Grade 5 | Only hybrid | 186 | 81 |
| Grade 5 | Changed instructional mode | 8,982 | 2,339 |
| Grade 5 | Not reported | 18,203 | 11,525 |
| Grade 8 | Only remote | 4,546 | 15,113 |
| Grade 8 | Only in person | 494 | 4 |
| Grade 8 | Only hybrid | 175 | 4 |
| Grade 8 | Changed instructional mode | 5,264 | 1,950 |
| Grade 8 | Not reported | 16,434 | 19,742 |

Table 7.16 *(continuation)*

|  |  |  |  |
| --- | --- | --- | --- |
| Grade Level | Instructional Mode | In-Person Assessment | Remote Assessment |
| Grade 10 | Only remote | 96 | 530 |
| Grade 10 | Only in person | 16 | 0 |
| Grade 10 | Only hybrid | 23 | 0 |
| Grade 10 | Changed instructional mode | 89 | 0 |
| Grade 10 | Not reported | 474 | 884 |
| Grade 11 | Only remote | 1,657 | 12,223 |
| Grade 11 | Only in person | 230 | 22 |
| Grade 11 | Only hybrid | 427 | 1 |
| Grade 11 | Changed instructional mode | 2,543 | 698 |
| Grade 11 | Not reported | 8,849 | 15,331 |
| Grade 12 | Only remote | 3,064 | 20,061 |
| Grade 12 | Only in person | 219 | 171 |
| Grade 12 | Only hybrid | 82 | 11 |
| Grade 12 | Changed instructional mode | 2,646 | 298 |
| Grade 12 | Not reported | 10,321 | 20,634 |
| High school | Only remote | 4,817 | 32,814 |
| High school | Only in person | 465 | 193 |
| High school | Only hybrid | 532 | 12 |
| High school | Changed instructional mode | 5,278 | 996 |
| High school | Not reported | 19,644 | 36,849 |

#### Analyses Providing Context

Analyses of testing issues and students’ testing experience were conducted to provide context for the interpretation of students’ scores. These analyses include the testing issues reported through the Security and Test Administration Incident Reporting System (STAIRS) in the Test Operations Management System, a comparison of the testing time for the remote and in-person test takers, and summaries of technology-related test disruptions (due to a browser crash, power shutdown, network loss, etc.).

STAIRS incidents are required when reporting several testing issues that may jeopardize the validity of score interpretations, including exposure of secure materials, student cheating, and inappropriate support provided to test takers (e.g., an adult coaching, a test administrator modifying responses, etc.). The rates of issues were well below a hundredth of 1 percent and relatively similar between the 2018–2019 and 2020–2021 administrations. Additionally, technical issues that may have impacted testing (e.g., administration errors or expired or accidentally submitted tests) were minimal and similar between the two administrations.

For those who had technology-related test disruptions that required a student to restart the test, most test takers had one restart; a small percentage (0.0–0.7%) of students had more than one restart. The number of test disruptions for the students testing remotely was slightly higher compared to the students testing in person.

#### Student Performance on Individual Test Questions

The performance of individual test questions, or items, was compared for in-person and remote test takers. Two types of item analyses were conducted: DIF and a comparison of IRT-based item parameter estimates.

##### Differential Item Functioning

DIF methodology was used to identify items that perform differently for in-person and remote test takers. Of the 200 items that were evaluated using DIF across all the tests, none were identified as exhibiting the most practically significant form of DIF between the students testing in person and remotely. Refer to section [*7.3 Differential Item Functioning Analyses*](#_Differential_Item_Functioning_1) for a description of DIF analysis, as it was applied to the results of the spring 2021 CAST administration.

##### Comparison of Item Response Theory Parameter Estimates

Comparisons of IRT-based item difficulties (i.e., *b*-parameter estimates) and item discriminations (i.e., *a*-parameter estimates) were conducted for grades five and eight and for high school. Table 7.17 presents the estimates for students testing in person and remotely, and shows minimal to no differences between the testing conditions.

Table 7.17 Average Item Parameter Estimates

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Remote a | In-Person a | Difference a | Remote b | In-Person b | Difference b |
| Grade 5 | 1.12 | 1.12 | 0.00 | 0.13 | 0.14 | -0.01 |
| Grade 8 | 1.06 | 1.05 | 0.01 | 0.64 | 0.64 | 0.00 |
| High school | 1.07 | 1.08 | -0.01 | 0.61 | 0.63 | -0.02 |

#### Student Performance on the Entire Test

Mean CAST scores for students within the in-person and remote groups were compared using statistical tests to evaluate whether differences were larger than differences that might be expected by chance alone. Total test scores for grades five and eight and for high school, were evaluated.

Differences were evaluated using an analysis of covariance model. The model included as covariates students’ prior Smarter Balanced mathematics scores,[[14]](#footnote-15) gender, ethnicity, Individuals with Disabilities Education Act status, Section 504 plan status, economic status, migrant education status, parental education, and whether students are from a charter school.

Considering only students who received remote instruction, differences were significant in grade eight and for high school; in each case, students who tested remotely outperformed students who tested in person. Table 7.18 presents the modeled difference in remote and in-person scale scores; the estimate is the mean difference in scale scores (i.e., remote minus in-person scale scores), controlling for the covariates included in the model.

Table 7.18 Modeled Difference in Remote and In-Person Scale Scores: Remote Instruction Only

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grade Level | Estimate | Standard Error | t | Pr > |t| |
| Grade 5 | -0.12 | 1.11 | -0.11 | 0.910 |
| Grade 8 | 3.20 | 1.14 | 2.79 | 0.010 |
| High school | 2.24 | 0.77 | 2.91 | 0.004 |

#### Summary of Reliability and Validity Analyses

The results of the reliability and validity analyses suggest no significant threats to the psychometric properties of the assessments that were administered remotely. Total test response times were compared for students testing remotely and in person. For grade eight and for high school, remote test takers consistently had longer median response times than in-person test takers. However, these differences did not suggest problematic test delivery conditions. No differences were observed in the reliability estimates for in-person and remote test-taker scores.

#### Limitations and Caveats for Data Interpretation

Differences related to the performance of students testing remotely or in person were carefully evaluated with a number of methods. There is some evidence to suggest that mode of instruction and test location may affect student performance. However, it is important to keep in mind that these results were obtained under nontraditional learning and assessment conditions due to the novel coronavirus disease 2019 pandemic. More evidence would be needed to further evaluate the feasibility of being able to generalize these comparability results to other test-taker populations.

### 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® version 3.51: Flexible multilevel and multidimensional item analysis and test scoring [Computer software]. Chapel Hill, NC: Vector Psychometric Group.

California Department of Education. (2019b)*.* *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. (2021). *CAASPP post-test guide: Technical information for student score reports for CAASPP LEA and test site coordinators and research specialists.* 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.

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., pp. 443–‍507). 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.

Embretson (Whitley), S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin,* *93*, 179–97.

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

Flack, V. F., Afifi, A. A., Lachenbruch, P. A., & Schouten, H. J. A. (1988). Sample size determinations for the two rater Kappa statistics. *Psychometrika,* *53*(3), 321–25.

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*, 347–60.

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.

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.

Landis, J. R., & Koch, G. G. (1977). The measurement of interrater agreement for categorical data. *Biometrics*, *33*, 159–74.

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.

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.

Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment*, *5*, 299–321.

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: 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 7.1

P value sub dich equals the fraction with the numerator the sum of X sub ic and the denominator N sub i end fraction.

#### Alternative Text for Equation 7.2

P value sub poly equals the fraction with the numerator the sum of X sub ij over j and the denominator N sub i times Max of X sub i end fraction.

#### Alternative Text for Equation 7.3

R sub polyreg equals the fraction with the numerator Beta hat times S sub tot and the denominator the square root of beta hat squared times s sub tot squared plus 1 I end fraction.

#### Alternative Text for Equation 7.4

Alpha sub MH equals the fraction with the numerator open bracket the sum from m of R sub rm times W sub fm divided by N sub tm close bracket and the denominator open bracket the sum from m of R sub fm times W sub rm divided by N sub tm close bracket end fraction.

#### Alternative Text for Equation 7.5

MH D - DIF equals negative 2.35 times the natural log of open bracket alpha sub MH close bracket.

#### Alternative Text for Equation 7.6

SMD equals the fraction with numerator the sum from m equals 1 to M of N sub fm times E sub f of Y from X equals m 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 r of Y from X equals m 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.

#### Alternative Text for Equation 7.7

P sub ih of theta sub j equals the numerator exp open parenthesis the sum from v equals 1 to h of D times a sub i of the quantity open parenthesis theta sub j minus b sub i plus d sub iv close parenthesis close parenthesis and denominator 1 plus the sum from c equals 1 to n sub i exp open parenthesis the sum from v equals 1 to c D times a sub i of the quantity open parenthesis theta sub j minus b sub i plus d sub iv close parenthesis close parenthesis, if score h equals 1, 2, …., n sub i.

P sub ih of theta sub j equals 1 divided by denominator 1 plus the sum from c equals 1 to n sub i exp open parenthesis the sum from v equals 1 to c D times a sub i of the quantity open parenthesis theta sub j minus b sub i plus d sub iv close parenthesis close parenthesis, if score h equals 0.

#### Alternative Text for Equation 7.8

Rho sub theta prime equals 1 minus M sub SEM squared sub theta divided by s squared sub theta.

#### Alternative Text for Equation 7.9

SEM sub theta equals square root of M sub open parenthesis of SE sub theta close parenthesis square.

#### Alternative Text for Equation 7.10

SEM sub scale score equals square root of M sub open parenthesis of CSEM sub scale score close parenthesis square.

#### Alternative Text for Equation 7.11

kappa equals the fraction with the numerator p sub obs minus p sub exp the denominator 1 minus p sub exp.

#### Alternative Text for Equation 7.12

P sub obs equals 1 divided by n times the sum from s equals 0 to m n sub ss.

#### Alternative Text for Equation 7.13

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.

#### Alternative Text for Equation 7.14

K sub st equals the fraction with numerator open parenthesis the sum from s equals zero to m the sum from t equals zero to m of w sub st times n sub st divided by n sub plus plus close parenthesis minus open parenthesis the sum from s equals zero to m the sum from t equals zero to m of w sub st times n sub splus times n sub plus t divided by n squared sub plus plus close parenthesis and the denominator 1 minus open parenthesis the sum from s equals zero to m the sum from t equals zero to m of w sub st times n sub splus times n sub plus t divided by n squared sub plus plus close parenthesis end fraction.

#### Alternative Text for Equation 7.15

W sub st equals 1 minus fraction with numerator open parenthesis s minus t close parenthesis squared and denominator m squared end fraction.

### Appendix 7.A: 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 7.A.1 Item Difficulty Distributions by Item Type

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Grade or Grade Level | 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 | 0 | 7 | 19 | 7 | 0 | 33 |
| Grade 5 | CR | 0 | 3 | 5 | 1 | 0 | 9 |
| Grade 5 | TEI | 1 | 5 | 6 | 8 | 1 | 21 |
| Grade 5 | Composite | 0 | 0 | 5 | 1 | 0 | 6 |
| Grade 8 | MC | 0 | 10 | 13 | 1 | 0 | 24 |
| Grade 8 | CR | 0 | 6 | 3 | 2 | 0 | 11 |
| Grade 8 | TEI | 1 | 14 | 14 | 0 | 0 | 29 |
| Grade 8 | Composite | 0 | 1 | 4 | 0 | 0 | 5 |
| High school—Grade 10 | MC | 1 | 8 | 10 | 7 | 0 | 26 |
| High school—Grade 10 | CR | 1 | 4 | 4 | 0 | 0 | 9 |
| High school—Grade 10 | TEI | 1 | 4 | 11 | 6 | 1 | 23 |
| High school—Grade 10 | Composite | 0 | 2 | 2 | 0 | 0 | 4 |
| High school—Grade 11 | MC | 1 | 7 | 13 | 5 | 0 | 26 |
| High school—Grade 11 | CR | 1 | 4 | 4 | 0 | 0 | 9 |
| High school—Grade 11 | TEI | 1 | 6 | 11 | 5 | 0 | 23 |
| High school—Grade 11 | Composite | 0 | 2 | 2 | 0 | 0 | 4 |
| High school—Grade 12 | MC | 2 | 8 | 11 | 5 | 0 | 26 |
| High school—Grade 12 | CR | 1 | 5 | 3 | 0 | 0 | 9 |
| High school—Grade 12 | TEI | 3 | 5 | 11 | 4 | 0 | 23 |
| High school—Grade 12 | Composite | 0 | 2 | 2 | 0 | 0 | 4 |
| High school—All grades | MC | 1 | 9 | 11 | 5 | 0 | 26 |
| High school—All grades | CR | 1 | 4 | 4 | 0 | 0 | 9 |
| High school—All grades | TEI | 2 | 6 | 10 | 5 | 0 | 23 |
| High school—All grades | Composite | 0 | 2 | 2 | 0 | 0 | 4 |

Table 7.A.2 Item Difficulty Distributions by Content Domain

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Grade or Grade Level | 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 | PS | 1 | 6 | 11 | 5 | 1 | 24 |
| Grade 5 | LS | 0 | 6 | 8 | 9 | 0 | 23 |
| Grade 5 | ESS | 0 | 3 | 16 | 3 | 0 | 22 |
| Grade 8 | PS | 0 | 11 | 12 | 1 | 0 | 24 |
| Grade 8 | LS | 1 | 10 | 13 | 1 | 0 | 25 |
| Grade 8 | ESS | 0 | 10 | 9 | 1 | 0 | 20 |
| High school—Grade 10 | PS | 0 | 5 | 9 | 3 | 0 | 17 |
| High school—Grade 10 | LS | 2 | 5 | 9 | 6 | 0 | 22 |
| High school—Grade 10 | ESS | 1 | 8 | 9 | 4 | 1 | 23 |
| High school—Grade 11 | PS | 0 | 4 | 10 | 3 | 0 | 17 |
| High school—Grade 11 | LS | 3 | 6 | 10 | 3 | 0 | 22 |
| High school—Grade 11 | ESS | 0 | 9 | 10 | 4 | 0 | 23 |
| High school—Grade 12 | PS | 0 | 6 | 8 | 3 | 0 | 17 |
| High school—Grade 12 | LS | 4 | 6 | 10 | 2 | 0 | 22 |
| High school—Grade 12 | ESS | 2 | 8 | 9 | 4 | 0 | 23 |
| High school—All grades | PS | 0 | 6 | 8 | 3 | 0 | 17 |
| High school—All grades | LS | 3 | 7 | 9 | 3 | 0 | 22 |
| High school—All grades | ESS | 1 | 8 | 10 | 4 | 0 | 23 |

**Note:** “PS” = Physical Sciences, “LS” = Life Sciences, “ESS” = Earth and Space Sciences.

### Appendix 7.B: 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 7.B.1 Item-Total Correlation Distributions by Item Type

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade or Grade Level | 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 | 1 | 4 | 6 | 22 | 33 |
| Grade 5 | CR | 0 | 0 | 0 | 0 | 1 | 8 | 9 |
| Grade 5 | TEI | 0 | 0 | 0 | 1 | 5 | 15 | 21 |
| Grade 5 | Composite | 0 | 0 | 0 | 1 | 1 | 4 | 6 |
| Grade 8 | MC | 0 | 1 | 2 | 4 | 3 | 14 | 24 |
| Grade 8 | CR | 0 | 0 | 0 | 0 | 1 | 10 | 11 |
| Grade 8 | TEI | 0 | 1 | 1 | 3 | 7 | 17 | 29 |
| Grade 8 | Composite | 0 | 0 | 0 | 0 | 1 | 4 | 5 |
| High school—Grade 10 | MC | 0 | 1 | 1 | 2 | 7 | 15 | 26 |
| High school—Grade 10 | CR | 0 | 0 | 0 | 0 | 0 | 9 | 9 |
| High school—Grade 10 | TEI | 0 | 1 | 0 | 2 | 3 | 17 | 23 |
| High school—Grade 10 | Composite | 0 | 0 | 0 | 0 | 1 | 3 | 4 |
| High school—Grade 11 | MC | 0 | 2 | 0 | 4 | 7 | 13 | 26 |
| High school—Grade 11 | CR | 0 | 0 | 0 | 0 | 0 | 9 | 9 |
| High school—Grade 11 | TEI | 0 | 1 | 0 | 2 | 6 | 14 | 23 |
| High school—Grade 11 | Composite | 0 | 0 | 0 | 1 | 0 | 3 | 4 |
| High school—Grade 12 | MC | 0 | 2 | 0 | 4 | 7 | 13 | 26 |
| High school—Grade 12 | CR | 0 | 0 | 0 | 0 | 0 | 9 | 9 |
| High school—Grade 12 | TEI | 0 | 1 | 0 | 3 | 4 | 15 | 23 |
| High school—Grade 12 | Composite | 0 | 0 | 0 | 1 | 0 | 3 | 4 |
| High school—All grades | MC | 0 | 2 | 0 | 4 | 7 | 13 | 26 |
| High school—All grades | CR | 0 | 0 | 0 | 0 | 0 | 9 | 9 |
| High school—All grades | TEI | 0 | 1 | 0 | 3 | 4 | 15 | 23 |
| High school—All grades | Composite | 0 | 0 | 0 | 1 | 0 | 3 | 4 |

Table 7.B.2 Item-Total Correlation Distributions by Content Domain

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade or Grade Level | 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 | PS | 0 | 0 | 1 | 4 | 4 | 15 | 24 |
| Grade 5 | LS | 0 | 0 | 0 | 1 | 2 | 20 | 23 |
| Grade 5 | ESS | 0 | 0 | 0 | 1 | 7 | 14 | 22 |
| Grade 8 | PS | 0 | 0 | 2 | 2 | 4 | 16 | 24 |
| Grade 8 | LS | 0 | 0 | 1 | 3 | 4 | 17 | 25 |
| Grade 8 | ESS | 0 | 2 | 0 | 2 | 4 | 12 | 20 |
| High school—Grade 10 | PS | 0 | 1 | 0 | 1 | 5 | 10 | 17 |
| High school—Grade 10 | LS | 0 | 1 | 0 | 1 | 4 | 16 | 22 |
| High school—Grade 10 | ESS | 0 | 0 | 1 | 2 | 2 | 18 | 23 |
| High school—Grade 11 | PS | 0 | 1 | 0 | 1 | 7 | 8 | 17 |
| High school—Grade 11 | LS | 0 | 1 | 0 | 3 | 5 | 13 | 22 |
| High school—Grade 11 | ESS | 0 | 1 | 0 | 3 | 1 | 18 | 23 |
| High school—Grade 12 | PS | 0 | 1 | 0 | 2 | 6 | 8 | 17 |
| High school—Grade 12 | LS | 0 | 1 | 0 | 3 | 4 | 14 | 22 |
| High school—Grade 12 | ESS | 0 | 1 | 0 | 3 | 1 | 18 | 23 |
| High school—All grades | PS | 0 | 1 | 0 | 2 | 6 | 8 | 17 |
| High school—All grades | LS | 0 | 1 | 0 | 3 | 4 | 14 | 22 |
| High school—All grades | ESS | 0 | 1 | 0 | 3 | 1 | 18 | 23 |

**Note:** “PS” = Physical Sciences, “LS” = Life Sciences, “ESS” = Earth and Space Sciences.

### Appendix 7.C: 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 7.C. 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 | 1 | 0 | 0 | 0 | 1 |
| 0.2 ≤ a < 0.4 | 5 | 1 | 3 | 2 | 11 |
| 0.4 ≤ a < 0.6 | 7 | 1 | 5 | 2 | 15 |
| 0.6 ≤ a < 0.8 | 10 | 2 | 6 | 1 | 19 |
| 0.8 ≤ a < 1.0 | 8 | 3 | 4 | 0 | 15 |
| 1.0 ≤ a < 1.2 | 0 | 2 | 3 | 1 | 6 |
| 1.2 ≤ a < 1.4 | 2 | 0 | 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 | 0.18 | 0.34 | 0.29 | 0.29 | 0.18 |
| Maximum | 1.24 | 1.14 | 1.04 | 1.02 | 1.24 |
| Mean | 0.67 | 0.75 | 0.69 | 0.54 | 0.67 |
| SD | 0.26 | 0.25 | 0.23 | 0.26 | 0.25 |
| **Number of Items** | **33** | **9** | **21** | **6** | **69** |

Table 7.C. 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 | 6 | 1 | 4 | 1 | 12 |
| 0.4 ≤ a < 0.6 | 6 | 4 | 8 | 3 | 21 |
| 0.6 ≤ a < 0.8 | 4 | 2 | 12 | 1 | 19 |
| 0.8 ≤ a < 1.0 | 7 | 3 | 4 | 0 | 14 |
| 1.0 ≤ a < 1.2 | 0 | 1 | 0 | 0 | 1 |
| 1.2 ≤ a < 1.4 | 0 | 0 | 0 | 0 | 0 |
| 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 | 0.09 | 0.30 | 0.12 | 0.39 | 0.09 |
| Maximum | 0.95 | 1.07 | 0.99 | 0.80 | 1.07 |
| Mean | 0.59 | 0.69 | 0.60 | 0.57 | 0.61 |
| SD | 0.26 | 0.24 | 0.20 | 0.15 | 0.22 |
| **Number of Items** | **24** | **11** | **29** | **5** | **69** |

Table 7.C. 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 | 2 | 0 | 1 | 0 | 3 |
| 0.2 ≤ a < 0.4 | 3 | 0 | 3 | 1 | 7 |
| 0.4 ≤ a < 0.6 | 10 | 1 | 6 | 3 | 20 |
| 0.6 ≤ a < 0.8 | 6 | 3 | 7 | 0 | 16 |
| 0.8 ≤ a < 1.0 | 3 | 2 | 2 | 0 | 7 |
| 1.0 ≤ a < 1.2 | 2 | 2 | 3 | 0 | 7 |
| 1.2 ≤ a < 1.4 | 0 | 1 | 0 | 0 | 1 |
| 1.4 ≤ a < 1.6 | 0 | 0 | 1 | 0 | 1 |
| 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 | 0.59 | 0.14 | 0.28 | 0.14 |
| Maximum | 1.19 | 1.30 | 1.56 | 0.53 | 1.56 |
| Mean | 0.59 | 0.89 | 0.70 | 0.45 | 0.66 |
| SD | 0.26 | 0.24 | 0.33 | 0.11 | 0.30 |
| **Number of Items** | **26** | **9** | **23** | **4** | **62** |

Table 7.C. Item Discrimination Parameter Distribution by Content Domain for Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| IRT-*a* Range | Life Sciences | Physical Sciences | Earth and Space Sciences | Number of Items |
| a < 0 | 0 | 0 | 0 | 0 |
| 0 ≤ a < 0.2 | 0 | 1 | 0 | 1 |
| 0.2 ≤ a < 0.4 | 3 | 5 | 3 | 11 |
| 0.4 ≤ a < 0.6 | 4 | 4 | 7 | 15 |
| 0.6 ≤ a < 0.8 | 5 | 8 | 6 | 19 |
| 0.8 ≤ a < 1.0 | 6 | 5 | 4 | 15 |
| 1.0 ≤ a < 1.2 | 4 | 1 | 1 | 6 |
| 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 | 0.27 | 0.18 | 0.34 | 0.18 |
| Maximum | 1.22 | 1.02 | 1.24 | 1.24 |
| Mean | 0.75 | 0.61 | 0.65 | 0.67 |
| SD | 0.25 | 0.24 | 0.24 | 0.25 |
| **Number of Items** | **23** | **24** | **22** | **69** |

Table 7.C. Item Discrimination Parameter Distribution by Content Domain for Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| IRT-*a* Range | Life Sciences | Physical Sciences | Earth and Space Sciences | Number of Items |
| a < 0 | 0 | 0 | 0 | 0 |
| 0 ≤ a < 0.2 | 0 | 0 | 2 | 2 |
| 0.2 ≤ a < 0.4 | 4 | 5 | 3 | 12 |
| 0.4 ≤ a < 0.6 | 9 | 5 | 7 | 21 |
| 0.6 ≤ a < 0.8 | 7 | 7 | 5 | 19 |
| 0.8 ≤ a < 1.0 | 4 | 7 | 3 | 14 |
| 1.0 ≤ a < 1.2 | 1 | 0 | 0 | 1 |
| 1.2 ≤ a < 1.4 | 0 | 0 | 0 | 0 |
| 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 | 0.22 | 0.21 | 0.09 | 0.09 |
| Maximum | 1.07 | 0.99 | 0.96 | 1.07 |
| Mean | 0.62 | 0.64 | 0.55 | 0.61 |
| SD | 0.20 | 0.24 | 0.23 | 0.22 |
| **Number of Items** | **25** | **24** | **20** | **69** |

Table 7.C. Item Discrimination Parameter Distribution by Content Domain for High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| IRT-*a* Range | Life Sciences | Physical Sciences | Earth and Space Sciences | Number of Items |
| a < 0 | 0 | 0 | 0 | 0 |
| 0 ≤ a < 0.2 | 1 | 1 | 1 | 3 |
| 0.2 ≤ a < 0.4 | 2 | 2 | 3 | 7 |
| 0.4 ≤ a < 0.6 | 8 | 8 | 4 | 20 |
| 0.6 ≤ a < 0.8 | 3 | 2 | 11 | 16 |
| 0.8 ≤ a < 1.0 | 3 | 3 | 1 | 7 |
| 1.0 ≤ a < 1.2 | 5 | 0 | 2 | 7 |
| 1.2 ≤ a < 1.4 | 0 | 0 | 1 | 1 |
| 1.4 ≤ a < 1.6 | 0 | 1 | 0 | 1 |
| 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 | 0.14 | 0.15 | 0.14 | 0.14 |
| Maximum | 1.16 | 1.56 | 1.30 | 1.56 |
| Mean | 0.69 | 0.64 | 0.66 | 0.66 |
| SD | 0.30 | 0.33 | 0.28 | 0.30 |
| **Number of Items** | **22** | **17** | **23** | **62** |

### Appendix 7.D: 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 7.D. 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 | 3 | 1 | 4 |
| −1.5 ≤ b < −1.0 | 2 | 0 | 3 | 0 | 5 |
| −1.0 ≤ b < −0.5 | 6 | 1 | 3 | 1 | 11 |
| −0.5 ≤ b < 0 | 11 | 3 | 3 | 3 | 20 |
| 0 ≤ b < 0.5 | 7 | 2 | 3 | 1 | 13 |
| 0.5 ≤ b < 1.0 | 3 | 1 | 4 | 0 | 8 |
| 1.0 ≤ b < 1.5 | 3 | 2 | 1 | 0 | 6 |
| 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 | -1.34 | -0.58 | -1.84 | -1.60 | -1.84 |
| Maximum | 1.82 | 1.42 | 1.75 | 0.03 | 1.82 |
| Mean | 0.00 | 0.34 | -0.23 | -0.50 | -0.07 |
| SD | 0.73 | 0.68 | 1.03 | 0.57 | 0.83 |
| **Number of Items** | **33** | **9** | **21** | **6** | **69** |

Table 7.D. 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 | 1 | 0 | 0 | 0 | 1 |
| −1.0 ≤ b < −0.5 | 1 | 1 | 2 | 1 | 5 |
| −0.5 ≤ b < 0 | 5 | 1 | 6 | 1 | 13 |
| 0 ≤ b < 0.5 | 7 | 5 | 6 | 2 | 20 |
| 0.5 ≤ b < 1.0 | 5 | 1 | 5 | 1 | 12 |
| 1.0 ≤ b < 1.5 | 3 | 2 | 5 | 0 | 10 |
| 1.5 ≤ b < 2.0 | 1 | 1 | 5 | 0 | 7 |
| 2.0 ≤ b < 2.5 | 1 | 0 | 0 | 0 | 1 |
| 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 | -0.67 | -0.82 | -0.56 | -1.15 |
| Maximum | 2.17 | 1.73 | 1.72 | 0.97 | 2.17 |
| Mean | 0.42 | 0.47 | 0.54 | 0.09 | 0.46 |
| SD | 0.76 | 0.69 | 0.78 | 0.63 | 0.74 |
| **Number of Items** | **24** | **11** | **29** | **5** | **69** |

Table 7.D. 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 | 0 | 0 | 0 |
| −1.5 ≤ b < −1.0 | 0 | 0 | 2 | 0 | 2 |
| −1.0 ≤ b < −0.5 | 5 | 0 | 2 | 0 | 7 |
| −0.5 ≤ b < 0 | 3 | 1 | 4 | 1 | 9 |
| 0 ≤ b < 0.5 | 6 | 3 | 5 | 1 | 15 |
| 0.5 ≤ b < 1.0 | 6 | 2 | 4 | 0 | 12 |
| 1.0 ≤ b < 1.5 | 2 | 2 | 2 | 1 | 7 |
| 1.5 ≤ b < 2.0 | 2 | 0 | 2 | 0 | 4 |
| 2.0 ≤ b < 2.5 | 0 | 1 | 1 | 1 | 3 |
| 2.5 ≤ b < 3.0 | 1 | 0 | 1 | 0 | 2 |
| 3.0 ≤ b < 3.5 | 0 | 0 | 0 | 0 | 0 |
| b ≥ 3.5 | 1 | 0 | 0 | 0 | 1 |
| Minimum | -0.72 | -0.02 | -1.12 | -0.25 | -1.12 |
| Maximum | 3.84 | 2.09 | 2.83 | 2.19 | 3.84 |
| Mean | 0.56 | 0.81 | 0.42 | 0.86 | 0.56 |
| SD | 1.08 | 0.71 | 1.01 | 1.03 | 1.00 |
| **Number of Items** | **26** | **9** | **23** | **4** | **62** |

Table 7.D. Item Difficulty Parameter Distribution by Content Domain for Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| IRT-*b* Range | Life Sciences | Physical Sciences | Earth and Space 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 | 2 | 2 | 0 | 4 |
| −1.5 ≤ b < −1.0 | 2 | 1 | 2 | 5 |
| −1.0 ≤ b < −0.5 | 5 | 4 | 2 | 11 |
| −0.5 ≤ b < 0 | 5 | 7 | 8 | 20 |
| 0 ≤ b < 0.5 | 3 | 3 | 7 | 13 |
| 0.5 ≤ b < 1.0 | 5 | 2 | 1 | 8 |
| 1.0 ≤ b < 1.5 | 1 | 4 | 1 | 6 |
| 1.5 ≤ b < 2.0 | 0 | 1 | 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 | -1.67 | -1.84 | -1.30 | -1.84 |
| Maximum | 1.27 | 1.75 | 1.82 | 1.82 |
| Mean | -0.23 | 0.02 | -0.01 | -0.07 |
| SD | 0.81 | 0.96 | 0.71 | 0.83 |
| **Number of Items** | **23** | **24** | **22** | **69** |

Table 7.D. Item Difficulty Parameter Distribution by Content Domain for Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| IRT-*b* Range | Life Sciences | Physical Sciences | Earth and Space 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 | 1 | 1 |
| −1.0 ≤ b < −0.5 | 4 | 0 | 1 | 5 |
| −0.5 ≤ b < 0 | 4 | 5 | 4 | 13 |
| 0 ≤ b < 0.5 | 6 | 11 | 3 | 20 |
| 0.5 ≤ b < 1.0 | 5 | 5 | 2 | 12 |
| 1.0 ≤ b < 1.5 | 4 | 1 | 5 | 10 |
| 1.5 ≤ b < 2.0 | 2 | 1 | 4 | 7 |
| 2.0 ≤ b < 2.5 | 0 | 1 | 0 | 1 |
| 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 | -0.82 | -0.45 | -1.15 | -1.15 |
| Maximum | 1.73 | 2.17 | 1.67 | 2.17 |
| Mean | 0.38 | 0.45 | 0.56 | 0.46 |
| SD | 0.76 | 0.62 | 0.88 | 0.74 |
| **Number of Items** | **25** | **24** | **20** | **69** |

Table 7.D. Item Difficulty Parameter Distribution by Content Domain for High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| IRT-*b* Range | Life Sciences | Physical Sciences | Earth and Space 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 | 2 | 2 |
| −1.0 ≤ b < −0.5 | 2 | 3 | 2 | 7 |
| −0.5 ≤ b < 0 | 5 | 0 | 4 | 9 |
| 0 ≤ b < 0.5 | 4 | 6 | 5 | 15 |
| 0.5 ≤ b < 1.0 | 4 | 5 | 3 | 12 |
| 1.0 ≤ b < 1.5 | 3 | 1 | 3 | 7 |
| 1.5 ≤ b < 2.0 | 2 | 2 | 0 | 4 |
| 2.0 ≤ b < 2.5 | 1 | 0 | 2 | 3 |
| 2.5 ≤ b < 3.0 | 1 | 0 | 1 | 2 |
| 3.0 ≤ b < 3.5 | 0 | 0 | 0 | 0 |
| b ≥ 3.5 | 0 | 0 | 1 | 1 |
| Minimum | -0.61 | -0.71 | -1.12 | -1.12 |
| Maximum | 2.68 | 1.64 | 3.84 | 3.84 |
| Mean | 0.60 | 0.44 | 0.63 | 0.56 |
| SD | 0.93 | 0.71 | 1.24 | 1.00 |
| **Number of Items** | **22** | **17** | **23** | **62** |

### Appendix 7.E: Testing Time Analyses

**Notes:**

* Testing time analyses were based on students who logged on to the test and whose total testing time at the test level did not equal zero.
* In table 7.E.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.
* Because testing time was recorded at the page level, items that were on a page with multiple items were excluded in the analysis in table 7.E.2.
* The following abbreviations apply in table 7.E.2:
* MCSS = multiple-choice single-select item
* CR = constructed-response item
* TE = technology-enhanced item
* Composite = composite item (an item type that includes multiple parts)
* HS = high school

Table 7.E. Testing Time (in Minutes) by Segment

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade or Grade Level | Segment | N | Mean | SD | Min | Max | 1st Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 99th Percentile |
| Grade 5 | Discrete Operational | 60,879 | 73.0 | 46.5 | 2.8 | 931.5 | 14.4 | 33.4 | 45.2 | 62.0 | 87.2 | 122.8 | 248.7 |
| Grade 5 | PT Operational | 60,879 | 18.6 | 14.6 | 0.6 | 342.6 | 1.9 | 5.7 | 10.2 | 15.6 | 22.8 | 33.3 | 72.4 |
| Grade 5 | Discrete Field Test | 30,525 | 19.7 | 15.2 | 0.7 | 346.9 | 2.1 | 6.9 | 11.2 | 16.4 | 23.8 | 34.9 | 75.4 |
| Grade 5 | PT Field Test | 30,354 | 9.6 | 8.1 | 0.0 | 171.7 | 0.8 | 2.7 | 5.0 | 7.9 | 11.7 | 17.3 | 40.2 |

Table 7.E.1 *(continuation one)*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade or Grade Level | Segment | N | Mean | SD | Min | Max | 1st Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 99th Percentile |
| Grade 8 | Discrete Operational | 69,585 | 80.4 | 53.2 | 3.0 | 1803.1 | 11.3 | 33.9 | 49.1 | 68.5 | 96.5 | 138.1 | 277.6 |
| Grade 8 | PT Operational | 69,585 | 14.5 | 12.0 | 0.0 | 433.0 | 1.3 | 3.9 | 7.5 | 12.2 | 18.1 | 26.5 | 58.8 |
| Grade 8 | Discrete Field Test | 34,672 | 17.9 | 14.9 | 0.1 | 456.2 | 1.2 | 4.3 | 9.4 | 15.0 | 22.2 | 32.8 | 73.5 |
| Grade 8 | PT Field Test | 34,913 | 8.6 | 8.0 | 0.0 | 235.9 | 0.6 | 2.0 | 4.1 | 7.0 | 10.7 | 16.0 | 38.9 |
| HS—Grade 10 | Discrete Operational | 2,176 | 61.5 | 41.0 | 3.8 | 468.2 | 9.1 | 28.2 | 40.2 | 53.3 | 71.8 | 96.6 | 223.5 |
| HS—Grade 10 | PT Operational | 2,176 | 13.3 | 9.8 | 0.7 | 123.7 | 1.3 | 4.5 | 7.5 | 11.5 | 16.2 | 23.3 | 50.9 |
| HS—Grade 10 | Discrete Field Test | 1,096 | 16.6 | 11.7 | 1.0 | 162.5 | 1.4 | 4.7 | 9.6 | 15.1 | 20.9 | 28.6 | 58.9 |
| HS—Grade 10 | PT Field Test | 1,080 | 7.7 | 6.7 | 0.4 | 75.9 | 0.6 | 2.0 | 3.6 | 6.3 | 9.4 | 14.7 | 34.4 |
| HS—Grade 11 | Discrete Operational | 41,848 | 55.5 | 35.0 | 1.9 | 675.5 | 5.9 | 21.6 | 34.4 | 49.4 | 68.7 | 92.9 | 182.9 |
| HS—Grade 11 | PT Operational | 41,848 | 11.7 | 9.0 | 0.1 | 303.7 | 1.0 | 3.0 | 6.1 | 10.1 | 15.0 | 21.2 | 43.9 |
| HS—Grade 11 | Discrete Field Test | 20,946 | 15.0 | 11.9 | 0.5 | 203.1 | 1.1 | 3.0 | 7.3 | 13.2 | 19.7 | 27.6 | 55.6 |
| HS—Grade 11 | PT Field Test | 20,902 | 6.5 | 5.8 | 0.0 | 128.2 | 0.5 | 1.3 | 2.9 | 5.3 | 8.5 | 12.5 | 26.6 |
| HS—Grade 12 | Discrete Operational | 59,189 | 50.6 | 30.2 | 1.3 | 759.4 | 6.0 | 20.3 | 32.1 | 45.8 | 62.5 | 83.8 | 154.8 |
| HS—Grade 12 | PT Operational | 59,189 | 10.4 | 7.9 | 0.4 | 202.3 | 1.0 | 2.7 | 5.4 | 9.0 | 13.4 | 18.9 | 37.3 |
| HS—Grade 12 | Discrete Field Test | 29,375 | 13.3 | 10.4 | 0.5 | 289.6 | 1.0 | 2.7 | 6.4 | 11.7 | 17.7 | 24.6 | 48.5 |
| HS—Grade 12 | PT Field Test | 29,814 | 5.7 | 5.0 | 0.0 | 149.4 | 0.5 | 1.2 | 2.5 | 4.6 | 7.5 | 10.9 | 23.6 |

Table 7.E.1 *(continuation two)*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade or Grade Level | Segment | N | Mean | SD | Min | Max | 1st Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 99th Percentile |
| HS—All Grades | Discrete Operational | 103,213 | 52.8 | 32.6 | 1.3 | 759.4 | 6.0 | 20.9 | 33.1 | 47.3 | 65.1 | 87.9 | 168.3 |
| HS—All Grades | PT Operational | 103,213 | 11.0 | 8.4 | 0.1 | 303.7 | 1.0 | 2.9 | 5.7 | 9.5 | 14.1 | 19.9 | 40.8 |
| HS—All Grades | Discrete Field Test | 51,417 | 14.1 | 11.1 | 0.5 | 289.6 | 1.0 | 2.8 | 6.8 | 12.4 | 18.6 | 25.9 | 51.8 |
| HS—All Grades | PT Field Test | 51,796 | 6.1 | 5.4 | 0.0 | 149.4 | 0.5 | 1.3 | 2.7 | 4.9 | 8.0 | 11.7 | 25.4 |

**Note:** Multiple-choice multiple-select (MCMS) items are included in the technology enhanced (TE) item category.

Table 7.E. Testing Time (in Minutes) by Item Type

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade or Grade Level | Item Type | N | Mean | SD | Min | Max | 1st Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 99th Percentile |
| Grade 5 | MCSS | 1,049,896 | 1.5 | 2.2 | 0.0 | 153.1 | 0.0 | 0.2 | 0.6 | 1.0 | 1.6 | 2.8 | 9.5 |
| Grade 5 | CR | 336,519 | 5.5 | 6.8 | 0.0 | 267.4 | 0.1 | 1.2 | 2.2 | 3.8 | 6.4 | 10.7 | 32.6 |
| Grade 5 | TE | 1,159,843 | 1.5 | 2.1 | 0.0 | 162.7 | 0.1 | 0.3 | 0.6 | 1.1 | 1.7 | 2.8 | 9.2 |
| Grade 5 | Composite | 411,123 | 2.3 | 3.0 | 0.0 | 273.8 | 0.2 | 0.6 | 1.0 | 1.6 | 2.6 | 4.1 | 13.0 |
| Grade 8 | MCSS | 832,820 | 1.5 | 2.3 | 0.0 | 436.7 | 0.0 | 0.1 | 0.5 | 1.0 | 1.7 | 3.0 | 10.2 |
| Grade 8 | CR | 406,858 | 5.5 | 7.0 | 0.0 | 361.9 | 0.1 | 1.2 | 2.2 | 3.8 | 6.4 | 10.9 | 33.0 |
| Grade 8 | TE | 1,493,073 | 1.7 | 2.5 | 0.0 | 252.5 | 0.1 | 0.3 | 0.7 | 1.2 | 2.0 | 3.3 | 10.9 |
| Grade 8 | Composite | 381,216 | 2.4 | 2.9 | 0.0 | 152.5 | 0.1 | 0.5 | 1.1 | 1.8 | 2.9 | 4.6 | 13.2 |
| HS—Grade 10 | MCSS | 40,974 | 1.4 | 1.9 | 0.0 | 117.4 | 0.0 | 0.2 | 0.6 | 1.0 | 1.6 | 2.6 | 7.8 |
| HS—Grade 10 | CR | 10,474 | 3.9 | 5.2 | 0.0 | 135.6 | 0.1 | 1.0 | 1.7 | 2.8 | 4.5 | 7.3 | 21.6 |
| HS—Grade 10 | TE | 39,559 | 1.4 | 1.7 | 0.0 | 41.4 | 0.1 | 0.2 | 0.6 | 1.0 | 1.7 | 2.6 | 7.4 |
| HS—Grade 10 | Composite | 9,338 | 2.1 | 2.3 | 0.0 | 55.2 | 0.1 | 0.5 | 1.0 | 1.7 | 2.5 | 4.0 | 10.9 |
| HS—Grade 11 | MCSS | 787,971 | 1.2 | 1.7 | 0.0 | 195.9 | 0.0 | 0.1 | 0.4 | 0.9 | 1.5 | 2.4 | 7.3 |
| HS—Grade 11 | CR | 201,207 | 3.6 | 4.4 | 0.0 | 272.5 | 0.1 | 0.7 | 1.5 | 2.6 | 4.3 | 7.0 | 20.0 |
| HS—Grade 11 | TE | 761,931 | 1.2 | 1.6 | 0.0 | 163.4 | 0.1 | 0.2 | 0.5 | 0.9 | 1.6 | 2.5 | 6.6 |
| HS—Grade 11 | Composite | 178,987 | 2.0 | 2.4 | 0.0 | 201.4 | 0.1 | 0.3 | 0.9 | 1.5 | 2.4 | 3.7 | 9.8 |
| HS—Grade 12 | MCSS | 1,110,598 | 1.1 | 1.5 | 0.0 | 82.5 | 0.0 | 0.1 | 0.4 | 0.8 | 1.4 | 2.2 | 6.0 |
| HS—Grade 12 | CR | 284,278 | 3.2 | 3.5 | 0.0 | 158.7 | 0.1 | 0.7 | 1.4 | 2.4 | 3.9 | 6.1 | 16.0 |
| HS—Grade 12 | TE | 1,073,044 | 1.1 | 1.4 | 0.0 | 103.4 | 0.1 | 0.2 | 0.4 | 0.8 | 1.5 | 2.3 | 5.6 |
| HS—Grade 12 | Composite | 252,349 | 1.8 | 2.1 | 0.0 | 181.2 | 0.1 | 0.3 | 0.8 | 1.4 | 2.3 | 3.5 | 8.7 |

Table 7.E.2 *(continuation)*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade or Grade Level | Item Type | N | Mean | SD | Min | Max | 1st Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 99th Percentile |
| HS—All Grades | MCSS | 1,939,543 | 1.2 | 1.6 | 0.0 | 195.9 | 0.0 | 0.1 | 0.4 | 0.9 | 1.5 | 2.3 | 6.6 |
| HS—All Grades | CR | 495,959 | 3.4 | 3.9 | 0.0 | 272.5 | 0.1 | 0.7 | 1.4 | 2.5 | 4.0 | 6.5 | 17.9 |
| HS—All Grades | TE | 1,874,534 | 1.2 | 1.5 | 0.0 | 163.4 | 0.1 | 0.2 | 0.4 | 0.9 | 1.5 | 2.4 | 6.1 |
| HS—All Grades | Composite | 440,674 | 1.9 | 2.3 | 0.0 | 201.4 | 0.1 | 0.3 | 0.8 | 1.5 | 2.4 | 3.6 | 9.2 |

### Appendix 7.F: Reliability Analysis

Table 7.F. Reliabilities and SEMs by Demographic Groups for Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Reliability | Theta Score SEM | Theta Score Variance |
| All students | 62,129 | 0.92 | 0.31 | 1.11 |
| Male | 31,523 | 0.92 | 0.31 | 1.19 |
| Female | 30,604 | 0.91 | 0.30 | 1.03 |
| Nonbinary | 2 | N/A | N/A | N/A |
| EL | 11,335 | 0.85 | 0.30 | 0.57 |
| English only | 38,791 | 0.91 | 0.31 | 1.08 |
| Reclassified fluent English proficient (RFEP) | 9,614 | 0.89 | 0.30 | 0.83 |
| Initial fluent English proficient (IFEP) | 2,348 | 0.88 | 0.36 | 1.06 |
| Adult English learner (ADEL) | 0 | N/A | N/A | N/A |
| To be determined | 12 | 0.94 | 0.37 | 2.40 |
| Economically disadvantaged | 34,083 | 0.90 | 0.29 | 0.83 |
| Not economically disadvantaged | 28,046 | 0.90 | 0.32 | 1.07 |
| American Indian or Alaska Native (All) | 503 | 0.90 | 0.29 | 0.83 |
| Asian (All) | 7,057 | 0.89 | 0.35 | 1.16 |
| Native Hawaiian or Other Pacific Islander (All) | 237 | 0.91 | 0.30 | 0.95 |
| Filipino (All) | 1,126 | 0.89 | 0.30 | 0.86 |
| Hispanic or Latino (All) | 30,222 | 0.90 | 0.29 | 0.82 |
| Black or African American (All) | 2,255 | 0.90 | 0.30 | 0.88 |
| White (All) | 17,133 | 0.90 | 0.31 | 0.99 |
| Two or more races (All) | 3,596 | 0.91 | 0.32 | 1.13 |
| Special education services | 7,038 | 0.90 | 0.31 | 1.02 |
| No special education services | 55,091 | 0.91 | 0.31 | 1.05 |
| Migrant education | 626 | 0.88 | 0.30 | 0.75 |
| Not migrant education | 61,503 | 0.92 | 0.31 | 1.11 |

Table 7.F.1 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Reliability | Theta Score SEM | Theta Score Variance |
| Military | 688 | 0.90 | 0.30 | 0.89 |
| Not military | 61,441 | 0.92 | 0.31 | 1.12 |
| Homeless | 1,463 | 0.89 | 0.30 | 0.81 |
| Not homeless | 60,666 | 0.92 | 0.31 | 1.11 |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 147 | 0.90 | 0.29 | 0.82 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 356 | 0.89 | 0.29 | 0.79 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 5,393 | 0.86 | 0.37 | 0.96 |
| Asian (Primary ethnicity—Economically disadvantaged) | 1,664 | 0.91 | 0.30 | 1.03 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 100 | 0.90 | 0.30 | 0.86 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 137 | 0.89 | 0.30 | 0.84 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 783 | 0.89 | 0.31 | 0.86 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 343 | 0.89 | 0.29 | 0.79 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 7,327 | 0.91 | 0.30 | 0.92 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 22,895 | 0.89 | 0.29 | 0.74 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 682 | 0.91 | 0.29 | 0.95 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 1,573 | 0.88 | 0.30 | 0.77 |
| White (Primary ethnicity—Not economically disadvantaged) | 11,314 | 0.89 | 0.32 | 0.91 |
| White (Primary ethnicity—Economically disadvantaged) | 5,819 | 0.91 | 0.29 | 0.91 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 2,300 | 0.89 | 0.33 | 1.02 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 1,296 | 0.91 | 0.30 | 0.96 |

Table 7.F. Reliabilities and SEMs by Demographic Groups for Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Reliability | Theta Score SEM | Theta Score Variance |
| All students | 74,823 | 0.91 | 0.33 | 1.15 |
| Male | 38,235 | 0.91 | 0.33 | 1.28 |
| Female | 36,565 | 0.90 | 0.32 | 1.00 |
| Nonbinary | 23 | 0.93 | 0.32 | 1.51 |
| EL | 10,521 | 0.71 | 0.37 | 0.47 |
| English only | 42,533 | 0.91 | 0.32 | 1.15 |
| RFEP | 18,436 | 0.89 | 0.31 | 0.86 |
| IFEP | 3,296 | 0.91 | 0.35 | 1.27 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 6 | N/A | N/A | N/A |
| Economically disadvantaged | 41,596 | 0.87 | 0.33 | 0.84 |
| Not economically disadvantaged | 33,227 | 0.91 | 0.32 | 1.20 |
| American Indian or Alaska Native (All) | 553 | 0.88 | 0.33 | 0.88 |
| Asian (All) | 7,722 | 0.91 | 0.34 | 1.24 |
| Native Hawaiian or Other Pacific Islander (All) | 268 | 0.88 | 0.31 | 0.76 |
| Filipino (All) | 1,803 | 0.90 | 0.30 | 0.89 |
| Hispanic or Latino (All) | 38,335 | 0.87 | 0.33 | 0.82 |
| Black or African American (All) | 2,717 | 0.86 | 0.34 | 0.83 |
| White (All) | 19,658 | 0.91 | 0.31 | 1.13 |
| Two or more races (All) | 3,767 | 0.92 | 0.32 | 1.21 |
| Special education services | 8,187 | 0.82 | 0.37 | 0.78 |
| No special education services | 66,636 | 0.91 | 0.32 | 1.09 |
| Migrant education | 668 | 0.84 | 0.34 | 0.73 |
| Not migrant education | 74,155 | 0.91 | 0.33 | 1.15 |
| Military | 1,048 | 0.91 | 0.31 | 1.01 |
| Not military | 73,775 | 0.91 | 0.33 | 1.15 |
| Homeless | 1,773 | 0.84 | 0.35 | 0.74 |
| Not homeless | 73,050 | 0.91 | 0.33 | 1.15 |

Table 7.F.2 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Reliability | Theta Score SEM | Theta Score Variance |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 182 | 0.90 | 0.31 | 0.93 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 371 | 0.86 | 0.34 | 0.81 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 5,717 | 0.88 | 0.35 | 1.04 |
| Asian (Primary ethnicity—Economically disadvantaged) | 2,005 | 0.91 | 0.32 | 1.18 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 119 | 0.89 | 0.30 | 0.84 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 149 | 0.85 | 0.32 | 0.64 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 1,230 | 0.90 | 0.30 | 0.89 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 573 | 0.89 | 0.30 | 0.81 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 9,273 | 0.90 | 0.31 | 0.95 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 29,062 | 0.85 | 0.34 | 0.74 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 868 | 0.88 | 0.31 | 0.82 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 1,849 | 0.83 | 0.36 | 0.76 |
| White (Primary ethnicity—Not economically disadvantaged) | 13,384 | 0.91 | 0.31 | 1.11 |
| White (Primary ethnicity—Economically disadvantaged) | 6,274 | 0.89 | 0.31 | 0.92 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 2,454 | 0.91 | 0.32 | 1.14 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 1,313 | 0.90 | 0.32 | 0.95 |

Table 7.F. Reliabilities and SEMs by Demographic Groups for Grade Ten

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Reliability | Theta Score SEM | Theta Score Variance |
| All students | 2,258 | 0.92 | 0.32 | 1.22 |
| Male | 1,124 | 0.92 | 0.32 | 1.33 |
| Female | 1,133 | 0.91 | 0.31 | 1.11 |
| Nonbinary | 1 | N/A | N/A | N/A |
| EL | 156 | 0.67 | 0.33 | 0.33 |
| English only | 1,444 | 0.91 | 0.31 | 1.08 |
| RFEP | 518 | 0.91 | 0.32 | 1.15 |
| IFEP | 140 | 0.90 | 0.37 | 1.45 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 0 | N/A | N/A | N/A |
| Economically disadvantaged | 992 | 0.87 | 0.30 | 0.73 |
| Not economically disadvantaged | 1,266 | 0.91 | 0.33 | 1.25 |
| American Indian or Alaska Native (All) | 22 | 0.85 | 0.30 | 0.62 |
| Asian (All) | 312 | 0.88 | 0.38 | 1.19 |
| Native Hawaiian or Other Pacific Islander (All) | 7 | N/A | N/A | N/A |
| Filipino (All) | 19 | 0.89 | 0.30 | 0.84 |
| Hispanic or Latino (All) | 890 | 0.87 | 0.30 | 0.69 |
| Black or African American (All) | 53 | 0.90 | 0.32 | 0.97 |
| White (All) | 830 | 0.90 | 0.31 | 0.92 |
| Two or more races (All) | 125 | 0.92 | 0.34 | 1.53 |
| Special education services | 226 | 0.87 | 0.33 | 0.87 |
| No special education services | 2,032 | 0.91 | 0.32 | 1.17 |
| Migrant education | 13 | 0.84 | 0.34 | 0.70 |
| Not migrant education | 2,245 | 0.92 | 0.32 | 1.21 |
| Military | 14 | 0.87 | 0.28 | 0.63 |
| Not military | 2,244 | 0.92 | 0.32 | 1.22 |
| Homeless | 28 | 0.77 | 0.31 | 0.41 |
| Not homeless | 2,230 | 0.92 | 0.32 | 1.22 |

Table 7.F.3 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Reliability | Theta Score SEM | Theta Score Variance |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 9 | N/A | N/A | N/A |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 13 | 0.77 | 0.30 | 0.39 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 281 | 0.85 | 0.39 | 1.01 |
| Asian (Primary ethnicity—Economically disadvantaged) | 31 | 0.92 | 0.36 | 1.72 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 6 | N/A | N/A | N/A |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 1 | N/A | N/A | N/A |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 16 | 0.85 | 0.29 | 0.59 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 3 | N/A | N/A | N/A |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 276 | 0.89 | 0.30 | 0.78 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 614 | 0.85 | 0.31 | 0.61 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 26 | 0.92 | 0.32 | 1.23 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 27 | 0.83 | 0.32 | 0.59 |
| White (Primary ethnicity—Not economically disadvantaged) | 566 | 0.89 | 0.31 | 0.93 |
| White (Primary ethnicity—Economically disadvantaged) | 264 | 0.88 | 0.29 | 0.73 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 86 | 0.92 | 0.36 | 1.56 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 39 | 0.90 | 0.30 | 0.95 |

Table 7.F. Reliabilities and SEMs by Demographic Groups for Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Reliability | Theta Score SEM | Theta Score Variance |
| All students | 43,694 | 0.91 | 0.32 | 1.11 |
| Male | 21,795 | 0.91 | 0.32 | 1.22 |
| Female | 21,880 | 0.90 | 0.31 | 0.99 |
| Nonbinary | 19 | 0.90 | 0.32 | 1.10 |
| EL | 3,402 | 0.74 | 0.35 | 0.48 |
| English only | 24,120 | 0.91 | 0.32 | 1.13 |
| RFEP | 14,315 | 0.89 | 0.31 | 0.90 |
| IFEP | 1,848 | 0.91 | 0.33 | 1.24 |
| ADEL | 0 | N/A | N/A | N/A |
| To be determined | 1 | N/A | N/A | N/A |
| Economically disadvantaged | 23,780 | 0.89 | 0.32 | 0.88 |
| Not economically disadvantaged | 19,914 | 0.92 | 0.32 | 1.21 |
| American Indian or Alaska Native (All) | 304 | 0.90 | 0.31 | 0.95 |
| Asian (All) | 4,439 | 0.91 | 0.34 | 1.30 |
| Native Hawaiian or Other Pacific Islander (All) | 174 | 0.90 | 0.32 | 1.02 |
| Filipino (All) | 1,316 | 0.90 | 0.31 | 0.97 |
| Hispanic or Latino (All) | 21,600 | 0.88 | 0.32 | 0.84 |
| Black or African American (All) | 1,401 | 0.89 | 0.33 | 0.96 |
| White (All) | 12,686 | 0.91 | 0.31 | 1.13 |
| Two or more races (All) | 1,774 | 0.91 | 0.32 | 1.18 |
| Special education services | 3,724 | 0.83 | 0.35 | 0.73 |
| No special education services | 39,970 | 0.91 | 0.31 | 1.07 |
| Migrant education | 266 | 0.86 | 0.31 | 0.68 |
| Not migrant education | 43,428 | 0.91 | 0.32 | 1.11 |
| Military | 645 | 0.92 | 0.32 | 1.25 |
| Not military | 43,049 | 0.91 | 0.32 | 1.10 |
| Homeless | 1,374 | 0.87 | 0.33 | 0.82 |
| Not homeless | 42,320 | 0.91 | 0.32 | 1.11 |

Table 7.F.4 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Reliability | Theta Score SEM | Theta Score Variance |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 139 | 0.89 | 0.30 | 0.85 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 165 | 0.90 | 0.32 | 1.01 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 2,699 | 0.90 | 0.35 | 1.24 |
| Asian (Primary ethnicity—Economically disadvantaged) | 1,740 | 0.91 | 0.32 | 1.12 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 81 | 0.91 | 0.32 | 1.22 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 93 | 0.86 | 0.32 | 0.73 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 864 | 0.90 | 0.32 | 0.96 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 452 | 0.89 | 0.31 | 0.91 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 5,512 | 0.90 | 0.31 | 0.97 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 16,088 | 0.87 | 0.32 | 0.78 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 519 | 0.91 | 0.32 | 1.10 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 882 | 0.87 | 0.33 | 0.84 |
| White (Primary ethnicity—Not economically disadvantaged) | 8,960 | 0.91 | 0.32 | 1.14 |
| White (Primary ethnicity—Economically disadvantaged) | 3,726 | 0.90 | 0.31 | 0.98 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 1,140 | 0.91 | 0.32 | 1.16 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 634 | 0.90 | 0.31 | 1.01 |

Table 7.F. Reliabilities and SEMs by Demographic Groups for Grade Twelve

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Reliability | Theta Score SEM | Theta Score Variance |
| All students | 61,607 | 0.90 | 0.32 | 1.02 |
| Male | 30,339 | 0.91 | 0.32 | 1.15 |
| Female | 31,246 | 0.89 | 0.31 | 0.90 |
| Nonbinary | 22 | 0.93 | 0.33 | 1.58 |
| EL | 5,086 | 0.73 | 0.35 | 0.46 |
| English only | 30,913 | 0.91 | 0.32 | 1.10 |
| RFEP | 22,461 | 0.88 | 0.31 | 0.81 |
| IFEP | 3,069 | 0.91 | 0.32 | 1.17 |
| ADEL | 69 | 0.70 | 0.33 | 0.36 |
| To be determined | 2 | N/A | N/A | N/A |
| Economically disadvantaged | 37,016 | 0.88 | 0.31 | 0.82 |
| Not economically disadvantaged | 24,591 | 0.91 | 0.32 | 1.16 |
| American Indian or Alaska Native (All) | 321 | 0.89 | 0.31 | 0.90 |
| Asian (All) | 5,240 | 0.91 | 0.33 | 1.25 |
| Native Hawaiian or Other Pacific Islander (All) | 241 | 0.87 | 0.31 | 0.74 |
| Filipino (All) | 1,913 | 0.89 | 0.31 | 0.89 |
| Hispanic or Latino (All) | 35,837 | 0.88 | 0.31 | 0.80 |
| Black or African American (All) | 2,419 | 0.88 | 0.33 | 0.87 |
| White (All) | 13,396 | 0.91 | 0.31 | 1.13 |
| Two or more races (All) | 2,240 | 0.91 | 0.32 | 1.14 |
| Special education services | 5,659 | 0.84 | 0.34 | 0.71 |
| No special education services | 55,948 | 0.90 | 0.31 | 1.00 |
| Migrant education | 474 | 0.86 | 0.32 | 0.70 |
| Not migrant education | 61,133 | 0.90 | 0.32 | 1.02 |
| Military | 956 | 0.91 | 0.31 | 1.04 |
| Not military | 60,651 | 0.90 | 0.32 | 1.02 |
| Homeless | 2,148 | 0.87 | 0.32 | 0.78 |
| Not homeless | 59,459 | 0.90 | 0.32 | 1.03 |

Table 7.F.5 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Reliability | Theta Score SEM | Theta Score Variance |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 116 | 0.90 | 0.31 | 0.92 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 205 | 0.89 | 0.31 | 0.87 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 3,147 | 0.91 | 0.34 | 1.22 |
| Asian (Primary ethnicity—Economically disadvantaged) | 2,093 | 0.91 | 0.31 | 1.09 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 109 | 0.87 | 0.30 | 0.69 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 132 | 0.86 | 0.32 | 0.76 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 1,195 | 0.89 | 0.30 | 0.88 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 718 | 0.88 | 0.33 | 0.87 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 8,196 | 0.90 | 0.31 | 0.93 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 27,641 | 0.87 | 0.31 | 0.74 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 815 | 0.90 | 0.32 | 0.98 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 1,604 | 0.86 | 0.33 | 0.76 |
| White (Primary ethnicity—Not economically disadvantaged) | 9,573 | 0.91 | 0.32 | 1.15 |
| White (Primary ethnicity—Economically disadvantaged) | 3,823 | 0.90 | 0.31 | 0.92 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 1,440 | 0.91 | 0.32 | 1.17 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 800 | 0.89 | 0.31 | 0.92 |

Table 7.F. Reliabilities and SEMs by Demographic Groups for High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Reliability | Theta Score SEM | Theta Score Variance |
| All students | 107,559 | 0.91 | 0.32 | 1.06 |
| Male | 53,258 | 0.91 | 0.32 | 1.18 |
| Female | 54,259 | 0.90 | 0.31 | 0.95 |
| Nonbinary | 42 | 0.92 | 0.33 | 1.35 |
| EL | 8,644 | 0.73 | 0.35 | 0.46 |
| English only | 56,477 | 0.91 | 0.32 | 1.11 |
| RFEP | 37,294 | 0.89 | 0.31 | 0.85 |
| IFEP | 5,057 | 0.91 | 0.32 | 1.23 |
| ADEL | 69 | 0.70 | 0.33 | 0.36 |
| To be determined | 3 | N/A | N/A | N/A |
| Economically disadvantaged | 61,788 | 0.88 | 0.32 | 0.84 |
| Not economically disadvantaged | 45,771 | 0.91 | 0.32 | 1.19 |
| American Indian or Alaska Native (All) | 647 | 0.90 | 0.31 | 0.91 |
| Asian (All) | 9,991 | 0.91 | 0.33 | 1.29 |
| Native Hawaiian or Other Pacific Islander (All) | 422 | 0.88 | 0.32 | 0.87 |
| Filipino (All) | 3,248 | 0.89 | 0.31 | 0.93 |
| Hispanic or Latino (All) | 58,327 | 0.88 | 0.31 | 0.81 |
| Black or African American (All) | 3,873 | 0.88 | 0.33 | 0.91 |
| White (All) | 26,912 | 0.91 | 0.31 | 1.12 |
| Two or more races (All) | 4,139 | 0.91 | 0.32 | 1.17 |
| Special education services | 9,609 | 0.84 | 0.34 | 0.72 |
| No special education services | 97,950 | 0.90 | 0.31 | 1.03 |
| Migrant education | 753 | 0.86 | 0.32 | 0.70 |
| Not migrant education | 106,806 | 0.91 | 0.32 | 1.07 |
| Military | 1,615 | 0.91 | 0.32 | 1.13 |
| Not military | 105,944 | 0.91 | 0.32 | 1.06 |
| Homeless | 3,550 | 0.87 | 0.32 | 0.79 |
| Not homeless | 104,009 | 0.91 | 0.32 | 1.07 |

Table 7.F.6 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Number Tested | Reliability | Theta Score SEM | Theta Score Variance |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 264 | 0.89 | 0.30 | 0.87 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 383 | 0.89 | 0.31 | 0.91 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 6,127 | 0.90 | 0.34 | 1.24 |
| Asian (Primary ethnicity—Economically disadvantaged) | 3,864 | 0.91 | 0.32 | 1.11 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 196 | 0.90 | 0.31 | 0.96 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 226 | 0.86 | 0.32 | 0.75 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 2,075 | 0.90 | 0.31 | 0.92 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 1,173 | 0.88 | 0.32 | 0.89 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 13,984 | 0.90 | 0.31 | 0.94 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 44,343 | 0.87 | 0.31 | 0.75 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 1,360 | 0.90 | 0.32 | 1.04 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 2,513 | 0.86 | 0.33 | 0.79 |
| White (Primary ethnicity—Not economically disadvantaged) | 19,099 | 0.91 | 0.32 | 1.14 |
| White (Primary ethnicity—Economically disadvantaged) | 7,813 | 0.90 | 0.31 | 0.94 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 2,666 | 0.91 | 0.32 | 1.19 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 1,473 | 0.90 | 0.31 | 0.96 |

Table 7.F. Reliabilities and SEMs by ELPAC Performance Levels

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grade or Grade Level | Group | Number of Students | Reliability | Theta Score SEM | Theta Score Variance |
| Grade 5 | Total Matched Cases | 11,167 | 0.85 | 0.30 | 0.57 |
| Grade 5 | Minimally Developed | 1,419 | 0.56 | 0.36 | 0.29 |
| Grade 5 | Somewhat Developed | 3,845 | 0.67 | 0.31 | 0.29 |
| Grade 5 | Moderately Developed | 3,980 | 0.76 | 0.27 | 0.30 |
| Grade 5 | Well Developed | 1,923 | 0.84 | 0.27 | 0.46 |
| Grade 8 | Total Matched Cases | 9,995 | 0.72 | 0.36 | 0.47 |
| Grade 8 | Minimally Developed | 1,564 | 0.44 | 0.43 | 0.33 |
| Grade 8 | Somewhat Developed | 2,895 | 0.52 | 0.39 | 0.32 |
| Grade 8 | Moderately Developed | 3,530 | 0.65 | 0.34 | 0.33 |
| Grade 8 | Well Developed | 2,006 | 0.79 | 0.30 | 0.44 |
| Grade 10 | Total Matched Cases | 152 | 0.66 | 0.33 | 0.31 |
| Grade 10 | Minimally Developed | 34 | 0.54 | 0.35 | 0.27 |
| Grade 10 | Somewhat Developed | 42 | 0.46 | 0.32 | 0.19 |
| Grade 10 | Moderately Developed | 51 | 0.50 | 0.31 | 0.19 |
| Grade 10 | Well Developed | 25 | 0.76 | 0.33 | 0.45 |
| Grade 11 | Total Matched Cases | 3,052 | 0.75 | 0.34 | 0.47 |
| Grade 11 | Minimally Developed | 596 | 0.50 | 0.39 | 0.30 |
| Grade 11 | Somewhat Developed | 1,076 | 0.63 | 0.36 | 0.34 |
| Grade 11 | Moderately Developed | 964 | 0.73 | 0.32 | 0.37 |
| Grade 11 | Well Developed | 416 | 0.81 | 0.30 | 0.45 |
| Grade 12 | Total Matched Cases | 4,374 | 0.73 | 0.35 | 0.46 |
| Grade 12 | Minimally Developed | 805 | 0.54 | 0.41 | 0.36 |
| Grade 12 | Somewhat Developed | 1,438 | 0.63 | 0.36 | 0.35 |
| Grade 12 | Moderately Developed | 1,440 | 0.69 | 0.32 | 0.33 |
| Grade 12 | Well Developed | 691 | 0.80 | 0.30 | 0.46 |

Table 7.F.7 *(continuation)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grade or Grade Level | Group | Number of Students | Reliability | Theta Score SEM | Theta Score Variance |
| High school | Total Matched Cases | 7,578 | 0.74 | 0.35 | 0.46 |
| High school | Minimally Developed | 1,435 | 0.53 | 0.40 | 0.33 |
| High school | Somewhat Developed | 2,556 | 0.63 | 0.36 | 0.35 |
| High school | Moderately Developed | 2,455 | 0.70 | 0.32 | 0.34 |
| High school | Well Developed | 1,132 | 0.80 | 0.30 | 0.46 |

**Note:** “Total Matched Cases” indicates students who took both the CAST and the ELPAC in 2020–2021. “Minimally Developed,” “Somewhat Developed,” “Moderately Developed,” and “Well Developed” indicate the overall English performance level attained by students, as defined by the ELPAC.

### Appendix 7.G: Analysis 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 0.17 in the *150–178* row in the *Standard Not Met* column in table 7.G.1 means that 17 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.
* 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 0.16 in the *150–‍178* row in the *Standard Not Met* column in table 7.G.2 means that 16 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.

Table 7.G. Grade Five: Decision Accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Placement Score | Standard Not Met | Standard Nearly Met | Standard Met | Standard Exceeded | Category Total |
| 150–178 | 0.17 | 0.04 | 0.00 | 0.00 | 0.22 |
| 179–213 | 0.03 | 0.42 | 0.03 | 0.00 | 0.48 |
| 214–230 | 0.00 | 0.04 | 0.14 | 0.02 | 0.20 |
| 231–250 | 0.00 | 0.00 | 0.02 | 0.08 | 0.10 |

All-forms average, estimated proportion correctly classified: total = 0.81 Standard Met and Exceeded = 0.93

Table 7.G. Grade Five: Decision Consistency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Placement Score | Standard Not Met | Standard Nearly Met | Standard Met | Standard Exceeded | Category Total |
| 150–178 | 0.16 | 0.05 | 0.00 | 0.00 | 0.22 |
| 179–213 | 0.05 | 0.39 | 0.05 | 0.00 | 0.48 |
| 214–230 | 0.00 | 0.05 | 0.11 | 0.03 | 0.20 |
| 231–250 | 0.00 | 0.00 | 0.03 | 0.07 | 0.10 |

Alternate form, estimated proportion consistently classified: total = 0.74 Standard Met and Exceeded = 0.90

Table 7.G. Grade Eight: Decision Accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Placement Score | Standard Not Met | Standard Nearly Met | Standard Met | Standard Exceeded | Category Total |
| 350–377 | 0.17 | 0.05 | 0.00 | 0.00 | 0.21 |
| 378–414 | 0.04 | 0.45 | 0.03 | 0.00 | 0.52 |
| 415–432 | 0.00 | 0.04 | 0.13 | 0.01 | 0.18 |
| 433–450 | 0.00 | 0.00 | 0.02 | 0.06 | 0.09 |

All-forms average, estimated proportion correctly classified: total = 0.81 Standard Met and Exceeded = 0.93

Table 7.G. Grade Eight: Decision Consistency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Placement Score | Standard Not Met | Standard Nearly Met | Standard Met | Standard Exceeded | Category Total |
| 350–377 | 0.15 | 0.06 | 0.00 | 0.00 | 0.21 |
| 378–414 | 0.06 | 0.41 | 0.05 | 0.00 | 0.52 |
| 415–432 | 0.00 | 0.05 | 0.11 | 0.03 | 0.18 |
| 433–450 | 0.00 | 0.00 | 0.03 | 0.06 | 0.09 |

Alternate form, estimated proportion consistently classified: total = 0.73 Standard Met and Exceeded = 0.90

Table 7.G. High School—Grade Ten: Decision Accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Placement Score | Standard Not Met | Standard Nearly Met | Standard Met | Standard Exceeded | Category Total |
| 550–575 | 0.07 | 0.04 | 0.00 | 0.00 | 0.10 |
| 576–614 | 0.03 | 0.48 | 0.04 | 0.00 | 0.55 |
| 615–635 | 0.00 | 0.03 | 0.19 | 0.01 | 0.23 |
| 636–650 | 0.00 | 0.00 | 0.03 | 0.09 | 0.12 |

All-forms average, estimated proportion correctly classified: total = 0.82 Standard Met and Exceeded = 0.92

Table 7.G. High School—Grade Ten: Decision Consistency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Placement Score | Standard Not Met | Standard Nearly Met | Standard Met | Standard Exceeded | Category Total |
| 550–575 | 0.06 | 0.04 | 0.00 | 0.00 | 0.10 |
| 576–614 | 0.04 | 0.44 | 0.06 | 0.00 | 0.55 |
| 615–635 | 0.00 | 0.05 | 0.16 | 0.02 | 0.23 |
| 636–650 | 0.00 | 0.00 | 0.03 | 0.08 | 0.12 |

Alternate form, estimated proportion consistently classified: total = 0.75 Standard Met and Exceeded = 0.89

Table 7.G. High School—Grade Eleven: Decision Accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Placement Score | Standard Not Met | Standard Nearly Met | Standard Met | Standard Exceeded | Category Total |
| 550–575 | 0.10 | 0.04 | 0.00 | 0.00 | 0.14 |
| 576–614 | 0.03 | 0.49 | 0.04 | 0.00 | 0.55 |
| 615–635 | 0.00 | 0.04 | 0.19 | 0.01 | 0.24 |
| 636–650 | 0.00 | 0.00 | 0.02 | 0.05 | 0.07 |

All-forms average, estimated proportion correctly classified: total = 0.82 Standard Met and Exceeded = 0.92

Table 7.G. High School—Grade Eleven: Decision Consistency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Placement Score | Standard Not Met | Standard Nearly Met | Standard Met | Standard Exceeded | Category Total |
| 550–575 | 0.09 | 0.04 | 0.00 | 0.00 | 0.14 |
| 576–614 | 0.04 | 0.45 | 0.05 | 0.00 | 0.55 |
| 615–635 | 0.00 | 0.05 | 0.16 | 0.03 | 0.24 |
| 636–650 | 0.00 | 0.00 | 0.02 | 0.05 | 0.07 |

Alternate form, estimated proportion consistently classified: total = 0.75 Standard Met and Exceeded = 0.89

Table 7.G. High School—Grade Twelve: Decision Accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Placement Score | Standard Not Met | Standard Nearly Met | Standard Met | Standard Exceeded | Category Total |
| 550–575 | 0.10 | 0.04 | 0.00 | 0.00 | 0.14 |
| 576–614 | 0.03 | 0.52 | 0.03 | 0.00 | 0.58 |
| 615–635 | 0.00 | 0.04 | 0.17 | 0.01 | 0.22 |
| 636–650 | 0.00 | 0.00 | 0.02 | 0.04 | 0.06 |

All-forms average, estimated proportion correctly classified: total = 0.83 Standard Met and Exceeded = 0.92

Table 7.G. High School—Grade Twelve: Decision Consistency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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 | 0.48 | 0.05 | 0.00 | 0.58 |
| 615–635 | 0.00 | 0.05 | 0.14 | 0.02 | 0.22 |
| 636–650 | 0.00 | 0.00 | 0.02 | 0.04 | 0.06 |

Alternate form, estimated proportion consistently classified: total = 0.76 Standard Met and Exceeded = 0.89

Table 7.G. High School: Decision Accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Placement Score | Standard Not Met | Standard Nearly Met | Standard Met | Standard Exceeded | Category Total |
| 550–575 | 0.10 | 0.04 | 0.00 | 0.00 | 0.14 |
| 576–614 | 0.03 | 0.51 | 0.03 | 0.00 | 0.57 |
| 615–635 | 0.00 | 0.04 | 0.17 | 0.01 | 0.23 |
| 636–650 | 0.00 | 0.00 | 0.02 | 0.04 | 0.06 |

All-forms average, estimated proportion correctly classified: total = 0.82 Standard Met and Exceeded = 0.92

Table 7.G. High School: Decision Consistency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Placement Score | Standard Not Met | Standard Nearly Met | Standard Met | Standard Exceeded | Category Total |
| 550–575 | 0.09 | 0.05 | 0.00 | 0.00 | 0.14 |
| 576–614 | 0.05 | 0.47 | 0.05 | 0.00 | 0.57 |
| 615–635 | 0.00 | 0.05 | 0.15 | 0.02 | 0.23 |
| 636–650 | 0.00 | 0.00 | 0.02 | 0.04 | 0.06 |

Alternate form, estimated proportion consistently classified: total = 0.75 Standard Met and Exceeded = 0.89

### Appendix 7.H: Correlations to Smarter Balanced Test Scores

Table 7.H. Correlations with Smarter Balanced ELA Test Scores for Grade Five

|  |  |  |
| --- | --- | --- |
| Group | Number Tested | Correlation |
| All students | 50,912 | 0.82 |
| Male | 25,799 | 0.83 |
| Female | 25,112 | 0.82 |
| Nonbinary | 1 | N/A |
| EL | 8,973 | 0.72 |
| English only | 32,769 | 0.81 |
| RFEP | 7,331 | 0.76 |
| IFEP | 1,807 | 0.80 |
| ADEL | 0 | N/A |
| To be determined | 7 | N/A |
| English proficiency unknown | 25 | 0.86 |
| Economically disadvantaged | 27,890 | 0.78 |
| Not economically disadvantaged | 23,022 | 0.81 |
| American Indian or Alaska Native (All) | 453 | 0.78 |
| Asian (All) | 4,678 | 0.84 |
| Native Hawaiian or Other Pacific Islander (All) | 204 | 0.81 |
| Filipino (All) | 864 | 0.77 |
| Hispanic or Latino (All) | 24,524 | 0.79 |
| Black or African American (All) | 1,900 | 0.78 |
| White (All) | 15,177 | 0.80 |
| Two or more races (All) | 3,112 | 0.81 |
| Special education services | 5,830 | 0.79 |
| No special education services | 45,082 | 0.81 |
| Migrant education | 434 | 0.76 |
| Not migrant education | 50,478 | 0.82 |
| Military | 587 | 0.77 |
| Not military | 50,325 | 0.82 |
| Homeless | 1,113 | 0.78 |
| Not homeless | 49,799 | 0.82 |

Table 7.H.1 *(continuation)*

|  |  |  |
| --- | --- | --- |
| Group | Number Tested | Correlation |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 125 | 0.81 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 328 | 0.75 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 3,394 | 0.79 |
| Asian (Primary ethnicity—Economically disadvantaged) | 1,284 | 0.83 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 82 | 0.79 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 122 | 0.80 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 611 | 0.76 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 253 | 0.78 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 6,164 | 0.80 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 18,360 | 0.77 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 578 | 0.79 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 1,322 | 0.75 |
| White (Primary ethnicity—Not economically disadvantaged) | 10,054 | 0.78 |
| White (Primary ethnicity—Economically disadvantaged) | 5,123 | 0.78 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 2,014 | 0.79 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 1,098 | 0.78 |

Table 7.H. Correlations with Smarter Balanced Mathematics Test Scores for Grade Five

|  |  |  |
| --- | --- | --- |
| Group | Number Tested | Correlation |
| All students | 50,952 | 0.80 |
| Male | 25,824 | 0.80 |
| Female | 25,126 | 0.79 |
| Nonbinary | 2 | N/A |
| EL | 9,028 | 0.65 |
| English only | 32,751 | 0.79 |
| RFEP | 7,326 | 0.75 |
| IFEP | 1,811 | 0.80 |
| ADEL | 0 | N/A |
| To be determined | 10 | N/A |
| English proficiency unknown | 26 | 0.86 |
| Economically disadvantaged | 27,891 | 0.74 |
| Not economically disadvantaged | 23,061 | 0.79 |
| American Indian or Alaska Native (All) | 453 | 0.76 |
| Asian (All) | 4,686 | 0.82 |
| Native Hawaiian or Other Pacific Islander (All) | 202 | 0.76 |
| Filipino (All) | 865 | 0.78 |
| Hispanic or Latino (All) | 24,547 | 0.74 |
| Black or African American (All) | 1,891 | 0.73 |
| White (All) | 15,186 | 0.77 |
| Two or more races (All) | 3,122 | 0.78 |
| Special education services | 5,817 | 0.74 |
| No special education services | 45,135 | 0.79 |
| Migrant education | 434 | 0.73 |
| Not migrant education | 50,518 | 0.80 |
| Military | 586 | 0.75 |
| Not military | 50,366 | 0.80 |
| Homeless | 1,121 | 0.70 |
| Not homeless | 49,831 | 0.80 |

Table 7.H.2 *(continuation)*

|  |  |  |
| --- | --- | --- |
| Group | Number Tested | Correlation |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 126 | 0.76 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 327 | 0.75 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 3,401 | 0.77 |
| Asian (Primary ethnicity—Economically disadvantaged) | 1,285 | 0.80 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 81 | 0.75 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 121 | 0.71 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 609 | 0.79 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 256 | 0.75 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 6,177 | 0.76 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 18,370 | 0.71 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 575 | 0.76 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 1,316 | 0.69 |
| White (Primary ethnicity—Not economically disadvantaged) | 10,069 | 0.75 |
| White (Primary ethnicity—Economically disadvantaged) | 5,117 | 0.74 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 2,023 | 0.77 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 1,099 | 0.72 |

Table 7.H. Correlations with Smarter Balanced ELA Test Scores for Grade Eight

|  |  |  |
| --- | --- | --- |
| Group | Number Tested | Correlation |
| All students | 57,275 | 0.80 |
| Male | 29,240 | 0.81 |
| Female | 28,014 | 0.79 |
| Nonbinary | 21 | 0.74 |
| EL | 7,644 | 0.60 |
| English only | 33,796 | 0.80 |
| RFEP | 13,308 | 0.72 |
| IFEP | 2,497 | 0.76 |
| ADEL | 0 | N/A |
| To be determined | 4 | N/A |
| English proficiency unknown | 26 | 0.81 |
| Economically disadvantaged | 31,521 | 0.75 |
| Not economically disadvantaged | 25,754 | 0.80 |
| American Indian or Alaska Native (All) | 473 | 0.76 |
| Asian (All) | 5,229 | 0.80 |
| Native Hawaiian or Other Pacific Islander (All) | 210 | 0.70 |
| Filipino (All) | 1,398 | 0.75 |
| Hispanic or Latino (All) | 28,641 | 0.75 |
| Black or African American (All) | 2,182 | 0.75 |
| White (All) | 16,116 | 0.79 |
| Two or more races (All) | 3,026 | 0.80 |
| Special education services | 6,296 | 0.72 |
| No special education services | 50,979 | 0.78 |
| Migrant education | 409 | 0.75 |
| Not migrant education | 56,866 | 0.80 |
| Military | 736 | 0.78 |
| Not military | 56,539 | 0.80 |
| Homeless | 1,170 | 0.74 |
| Not homeless | 56,105 | 0.80 |

Table 7.H.3 *(continuation)*

|  |  |  |
| --- | --- | --- |
| Group | Number Tested | Correlation |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 154 | 0.80 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 319 | 0.73 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 3,692 | 0.74 |
| Asian (Primary ethnicity—Economically disadvantaged) | 1,537 | 0.82 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 89 | 0.76 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 121 | 0.61 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 955 | 0.74 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 443 | 0.74 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 7,293 | 0.78 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 21,348 | 0.72 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 723 | 0.74 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 1,459 | 0.73 |
| White (Primary ethnicity—Not economically disadvantaged) | 10,889 | 0.78 |
| White (Primary ethnicity—Economically disadvantaged) | 5,227 | 0.77 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 1,959 | 0.78 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 1,067 | 0.77 |

Table 7.H. Correlations with Smarter Balanced Mathematics Test Scores for Grade Eight

|  |  |  |
| --- | --- | --- |
| Group | Number Tested | Correlation |
| All students | 56,856 | 0.79 |
| Male | 29,016 | 0.80 |
| Female | 27,818 | 0.78 |
| Nonbinary | 22 | 0.66 |
| EL | 7,608 | 0.57 |
| English only | 33,513 | 0.79 |
| RFEP | 13,232 | 0.73 |
| IFEP | 2,472 | 0.81 |
| ADEL | 0 | N/A |
| To be determined | 5 | N/A |
| English proficiency unknown | 26 | 0.84 |
| Economically disadvantaged | 31,300 | 0.72 |
| Not economically disadvantaged | 25,556 | 0.80 |
| American Indian or Alaska Native (All) | 451 | 0.75 |
| Asian (All) | 5,244 | 0.81 |
| Native Hawaiian or Other Pacific Islander (All) | 210 | 0.73 |
| Filipino (All) | 1,394 | 0.77 |
| Hispanic or Latino (All) | 28,383 | 0.72 |
| Black or African American (All) | 2,146 | 0.72 |
| White (All) | 16,012 | 0.78 |
| Two or more races (All) | 3,016 | 0.80 |
| Special education services | 6,273 | 0.68 |
| No special education services | 50,583 | 0.78 |
| Migrant education | 421 | 0.64 |
| Not migrant education | 56,435 | 0.79 |
| Military | 722 | 0.76 |
| Not military | 56,134 | 0.79 |
| Homeless | 1,159 | 0.69 |
| Not homeless | 55,697 | 0.79 |

Table 7.H.4 *(continuation)*

|  |  |  |
| --- | --- | --- |
| Group | Number Tested | Correlation |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 144 | 0.76 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 307 | 0.73 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 3,700 | 0.77 |
| Asian (Primary ethnicity—Economically disadvantaged) | 1,544 | 0.80 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 90 | 0.77 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 120 | 0.63 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 955 | 0.78 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 439 | 0.73 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 7,174 | 0.76 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 21,209 | 0.69 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 708 | 0.74 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 1,438 | 0.68 |
| White (Primary ethnicity—Not economically disadvantaged) | 10,836 | 0.77 |
| White (Primary ethnicity—Economically disadvantaged) | 5,176 | 0.75 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 1,949 | 0.78 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 1,067 | 0.76 |

Table 7.H. Correlations with Smarter Balanced ELA Test Scores for Grade Eleven

|  |  |  |
| --- | --- | --- |
| Group | Number Tested | Correlation |
| All students | 38,225 | 0.78 |
| Male | 18,866 | 0.79 |
| Female | 19,343 | 0.78 |
| Nonbinary | 16 | 0.75 |
| EL | 2,802 | 0.60 |
| English only | 21,096 | 0.77 |
| RFEP | 12,707 | 0.74 |
| IFEP | 1,611 | 0.76 |
| ADEL | 0 | N/A |
| To be determined | 1 | N/A |
| English proficiency unknown | 8 | N/A |
| Economically disadvantaged | 20,851 | 0.75 |
| Not economically disadvantaged | 17,374 | 0.78 |
| American Indian or Alaska Native (All) | 274 | 0.77 |
| Asian (All) | 3,985 | 0.79 |
| Native Hawaiian or Other Pacific Islander (All) | 157 | 0.75 |
| Filipino (All) | 1,234 | 0.75 |
| Hispanic or Latino (All) | 18,702 | 0.75 |
| Black or African American (All) | 1,215 | 0.75 |
| White (All) | 11,095 | 0.76 |
| Two or more races (All) | 1,563 | 0.78 |
| Special education services | 3,089 | 0.70 |
| No special education services | 35,136 | 0.77 |
| Migrant education | 242 | 0.75 |
| Not migrant education | 37,983 | 0.78 |
| Military | 463 | 0.81 |
| Not military | 37,762 | 0.78 |
| Homeless | 1,187 | 0.72 |
| Not homeless | 37,038 | 0.78 |

Table 7.H.5 *(continuation)*

|  |  |  |
| --- | --- | --- |
| Group | Number Tested | Correlation |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 128 | 0.79 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 146 | 0.75 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 2,365 | 0.77 |
| Asian (Primary ethnicity—Economically disadvantaged) | 1,620 | 0.78 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 77 | 0.75 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 80 | 0.73 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 811 | 0.73 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 423 | 0.76 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 4,821 | 0.77 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 13,881 | 0.73 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 463 | 0.76 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 752 | 0.73 |
| White (Primary ethnicity—Not economically disadvantaged) | 7,724 | 0.76 |
| White (Primary ethnicity—Economically disadvantaged) | 3,371 | 0.74 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 985 | 0.78 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 578 | 0.76 |

Table 7.H. Correlations with Smarter Balanced Mathematics Test Scores for Grade Eleven

|  |  |  |
| --- | --- | --- |
| Group | Number Tested | Correlation |
| All students | 37,758 | 0.74 |
| Male | 18,634 | 0.74 |
| Female | 19,107 | 0.73 |
| Nonbinary | 17 | 0.80 |
| EL | 2,809 | 0.45 |
| English only | 20,710 | 0.74 |
| RFEP | 12,631 | 0.70 |
| IFEP | 1,599 | 0.73 |
| ADEL | 0 | N/A |
| To be determined | 1 | N/A |
| English proficiency unknown | 8 | N/A |
| Economically disadvantaged | 20,684 | 0.68 |
| Not economically disadvantaged | 17,074 | 0.75 |
| American Indian or Alaska Native (All) | 260 | 0.71 |
| Asian (All) | 3,992 | 0.77 |
| Native Hawaiian or Other Pacific Islander (All) | 156 | 0.75 |
| Filipino (All) | 1,227 | 0.74 |
| Hispanic or Latino (All) | 18,554 | 0.67 |
| Black or African American (All) | 1,206 | 0.65 |
| White (All) | 10,831 | 0.73 |
| Two or more races (All) | 1,532 | 0.74 |
| Special education services | 3,046 | 0.59 |
| No special education services | 34,712 | 0.72 |
| Migrant education | 239 | 0.66 |
| Not migrant education | 37,519 | 0.74 |
| Military | 455 | 0.78 |
| Not military | 37,303 | 0.74 |
| Homeless | 1,183 | 0.67 |
| Not homeless | 36,575 | 0.74 |

Table 7.H.6 *(continuation)*

|  |  |  |
| --- | --- | --- |
| Group | Number Tested | Correlation |
| American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged) | 123 | 0.73 |
| American Indian or Alaska Native (Primary ethnicity—Economically disadvantaged) | 137 | 0.70 |
| Asian (Primary ethnicity—Not economically disadvantaged) | 2,369 | 0.74 |
| Asian (Primary ethnicity—Economically disadvantaged) | 1,623 | 0.76 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged) | 76 | 0.77 |
| Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged) | 80 | 0.69 |
| Filipino (Primary ethnicity—Not economically disadvantaged) | 803 | 0.73 |
| Filipino (Primary ethnicity—Economically disadvantaged) | 424 | 0.73 |
| Hispanic or Latino (Primary ethnicity—Not economically disadvantaged) | 4,773 | 0.71 |
| Hispanic or Latino (Primary ethnicity—Economically disadvantaged) | 13,781 | 0.64 |
| Black or African American (Primary ethnicity—Not economically disadvantaged) | 454 | 0.70 |
| Black or African American (Primary ethnicity—Economically disadvantaged) | 752 | 0.60 |
| White (Primary ethnicity—Not economically disadvantaged) | 7,514 | 0.73 |
| White (Primary ethnicity—Economically disadvantaged) | 3,317 | 0.68 |
| Two or more races (Primary ethnicity—Not economically disadvantaged) | 962 | 0.76 |
| Two or more races (Primary ethnicity—Economically disadvantaged) | 570 | 0.67 |

## Quality Control Procedures

The California Department of Education (CDE) and ETS implemented rigorous quality control procedures throughout the test development, administration, scoring, analyses, and reporting processes for the California Science Test (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 Test Materials

Brief descriptions of other types of materials used for and during testing appear in the following subsections.

#### Developing Computer-based Assessments

The steps taken to develop and ensure the quality of the computer-based assessments are described in [*Chapter 3: Item Development*](#_Item_Development_1).

#### Test Administration Instructions

For the computer-based CAST, content was incorporated to fit the California Assessment of Student Performance and Progress (CAASPP) System specifications used to administer Smarter Balanced Summative Assessments. ETS staff consulted with internal subject matter experts and conducted validation checks to verify that test instructions accurately matched the testing processes. Copy editors and content editors reviewed each document for spelling, grammar, accuracy, and adherence to CDE style and usage requirements as well as the CDE accessibility standards.

For the paper–pencil tests (PPTs) that were provided upon approval from the CDE, ETS worked with the CDE to develop the scripts of instructions that were used by test administrators.

All CAASPP documents were approved by the CDE before they could be published to the CAASPP website. Only nonsecure documents were posted to this website.

##### 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.

#### Collecting Test Materials

##### Computer-based Assessments

During the 2020–2021 CAST administration, there were no collectable materials associated with computer-based testing.

##### Paper–Pencil Forms

Once the PPTs were administered at test sites whose local educational agencies (LEAs) had received prior approval from the CDE, LEAs were instructed to enter student responses into the CAASPP Data Entry Interface (DEI) before returning all scorable and nonscorable materials to ETS within five working days after the last day of each test administration period. The LEAs packed all materials into cartons, applied the labels, and then numbered the cartons prior to returning the materials to the processing center by means of their assigned carrier.

#### Processing Test Materials

##### Computer-based Assessments

Computer-based tests that were 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.

Test responses were sent for human scoring, and the reader’s ratings were delivered to ETS scoring systems for merging with machine-scored items, final scoring, and scoring quality checks.

##### Paper–Pencil Forms

Upon receipt of the test materials, ETS personnel examined each shipment of materials by comparing the serial numbers of returned materials with those on actual documents received. LEAs were contacted by phone if there were any missing materials.

### Quality Control of Item Development

ETS’ goal is to provide the best standards-based and innovative items for the CAST. Items developed for the CAST were subject to an extensive item review process. The item writers responsible for developing CAST items and performance tasks (PTs) were trained in CAASPP and ETS policies on quality control of item content, sensitivity, and bias guidelines, as well as guidelines for accessibility, to ensure that the items allow the widest possible range of students to demonstrate their content knowledge.

Once a written item was accepted for authoring—that is, once it was entered into ETS’ item bank and formatted for use in an assessment—ETS employed a series of internal and external reviews. These reviews used established criteria and specifications to judge the quality of item 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 and PTs 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).

During typical administrations of the CAST, when sufficient student response data on each item became available (2,000 responses for non–constructed-response [CR] items and 1,800 for CR items), ETS Psychometric Analysis & Research (PAR) staff conduct item analyses and a key check to examine whether the items perform as expected. ETS psychometric staff conduct a thorough evaluation of all item statistics using the statistical criteria described in subsection [*7.2.6 Summary of Classical Item Analyses Flagging Criteria*](#_Summary_of_Classical) to flag items that are potentially problematic because of poor item performance, content issues, item bias, or accessibility challenges. Flagged items are then reviewed by ETS Assessment and Learning Technology Research & Development (ALTRD) staff to determine whether issues exist.

### Quality Control of Test Form Development

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 ALTRD 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 answers were correct;
* items were scored correctly in the item bank and incorrect answers were scored as incorrect;
* all items assessed the standard;
* all content in the item was correct;
* all items met the statistical criteria;
* 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 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.

### Quality Control of Test Administration

The quality of test administration for the CAST, administered as part of the CAASPP System, was monitored and controlled through several strategies. A fully staffed support center, the California Technical Assistance Center (CalTAC), supported all local educational agencies (LEAs) in the administration of CAASPP assessments. In addition to providing guidance and answering questions, CalTAC regularly conducted outreach campaigns on particular administration topics to ensure all LEAs understood correct test administration procedures. CalTAC was guided by a core group of LEA outreach and advocacy staff that managed communications to LEAs; provided regional and web-based trainings; and hosted a website, [the](https://www.caaspp.org/) CAASPP 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 CAASPP assessments, including the CAST. 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 OTI developed and applied in quality control are described in subsection [*5.6.1 ETS’ Office of Testing Integrity*](#_ETS’s_Office_of).

### Quality Control of Scoring

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 a test 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 ALTRD staff reviewed and verified the keys and scoring rubrics for each item. Then, these keys and rubrics 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.

#### Quality Control of Machine-Scoring Procedures

To ensure valid item-level scoring for the CAST, quality control procedures were employed by Cambium Assessment, Inc., the CAASPP subcontractor responsible for providing the test delivery system (TDS) and scoring machine-scorable items. A real-time, quality-monitoring component was built into the TDS. After a test 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 test 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.

#### Quality Control of Human Scoring

For human scoring, ETS employed multiple quality controls including

* raters being required to successfully pass calibration, described in subsection [*6.1.1.2.6 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[*6.1.1.2 Human Scoring*](#_Human_Scoring) for the topics; refer to subsections [*6.1.1.2.8 Scoring Monitoring and Quality Management*](#_Scoring_Monitoring_and), and [*6.1.1.2.10 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 2020–2021 administration of the CAST, to ensure the quality of artificial intelligence (AI) scoring, ETS maintained a QA system where a random sample of human ratings were also obtained and used for rater agreement analyses. All AI-scored items were subject to a second scoring by human raters. With lower than expected volumes in 2020–2021, all items may not have met a suitable sample size for analysis. Refer to subsection [*6.1.1.2.8 Scoring Monitoring and Quality Management*](#_Scoring_Monitoring_and) for more information regarding the scoring process.

#### Enterprise Score Key Management System Processing

Prior to the test administration, ETS ALTRD staff reviewed and verified the keys and scoring rubrics for all items. Then, these keys and rubrics were provided to CAI for its machine‑scoring implementation. After CAI finished machine-scoring, those scores and responses were delivered to ETS. CAI quality control of the machine-scoring procedure is described in subsection [*8.5.1. Quality Control of Machine-Scoring Procedures*](#_Quality_Control_of_2)*.*

ETS’ Centralized Repository Distribution System and Enterprise Service Bus departments collected and parsed .xml files that contained student response data from CAI. ETS’ Enterprise Score Key Management (eSKM) system collected and calculated individual students’ overall scores (total raw scores) and generated student scores in the approved statistical extract format. These data extracts were sent to ETS’ Data Quality Services for data validation. Following successful validation, the student response statistical extracts were made available to the psychometric team.

ETS developed two parallel scoring systems to produce and verify student scores:

1. The eSKM scoring system received an individual student’s item scores and item responses from CAI and calculated individual student scores for ETS’ reporting systems.
2. The ETS PAR team computed individual student scores based on item scores delivered by CAI.

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 to ensure the quality and accuracy of scoring and to support the transfer of scores into the Test Operations Management System (TOMS), the database of the student records scoring system.

### Quality Control of Psychometric Processes

#### Development of Scoring Specifications

ETS scoring specifications for the CAST were completed, reviewed, approved, and checked in advance of the receipt of student response data. Before psychometric analysis, PAR developed a psychometric analysis plan and road map, describing each step of psychometric analyses, procedures, and schedules. This plan was submitted to the CDE for review and approval. After that, psychometric specifications were developed for ETS data analysts conducting all analyses. Psychometric specifications contained detailed scoring procedures as well as the procedures for determining whether a student attempted a test and whether that student’s response data should be included in the statistical analyses and calculations for computing summary data.

#### Psychometric Analyses

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 performed on the CAST data. 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 ALTRD staff for review; their comments were then reviewed by the psychometricians before the review by the CDE. The ETS ALTRD 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 fit the data. ETS then presented and reviewed the calibration results with the CDE for approval.

### Quality Control of Reporting

To ensure the quality of the 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 ETS’ Data Quality Services and Center of Reporting & Scoring Services teams, as well as running of all the Student Score Reports through ETS’ patented Quality Control Interrogator software, which compares elements of the student score report 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 California Educator Reporting System as well as via the LEA’s 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’s 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 the accuracy of 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 grades, 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 the CDE’s approval of the reports generated for the quality control LEAs, ETS continued with the report production. The quality control LEAs comprised CDE-selected LEAs to validate a subset of LEAs that contained key reporting characteristics (e.g., academic achievement) and demographics of the state. The reviewed set of reports incorporated CDE-selected LEAs and provided the final check prior to generating the reports and making them electronically available for download from 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 tests and answer documents that, for example, identified students who tested but responded to no items, who were not tested because of parent/guardian request, or who 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 subsection [*6.3.2 Special Cases*](#_Special_Cases).

### 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 was to verify that all systems, processes, and resources were ready for the operational administration. ETS employed a number of approaches 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 mirrored the final production environment, for operational testing and UAT. Final approval by the CDE triggered the final deployment of the system.

#### Computer-based Assessments

ETS employed a number of strategies to verify ongoing systems performance, including monitoring of system availability and computer-based 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 mirrored the final production environment, for operational testing and UAT. Final approval by the CDE triggered the final deployment of the system.

#### Paper–Pencil Tests

The Data Entry Interface (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.

### Reference

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 2020–2021 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 (CDE). In 2020–‍2021, the CDE used similar questions to those used in 2018–2019 to gather information about how the science content on the CAST compared to the science content presented in the classroom.

### Student Survey Questions

The survey questionnaire was administered after students completed the CAST. There were four survey questions:

1. Did you learn about the topics on the test in your science classes?
2. Were any questions on the test different from the types of questions you see in science classes?
3. How hard were questions on this test compared to questions you see in science classes?
4. Do you think you will be enrolling in any more science classes in high school?

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 9.A](#_Appendix_9.A:_Results) 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 9.A.1 through table 9.A.3 show the survey results for the three survey questions for grade five students. Table 9.A.4 through table 9.A.6 show the survey results for grade eight students. Table 9.A.7 through table 9.A.21 show the survey results for the survey questions by grade level and for high school students overall. Table 9.A.21 shows results only for grades ten and eleven students, since the fourth survey question does not apply to grade twelve students.

For the first survey question, “Did you learn about the topics on the test in your science classes?,” at least half of the students agreed that most or all of the topics were taught in their science classes. Specifically, this was endorsed by 66 percent of grade five students, 67 percent of grade eight students, and 51 percent of high school students. Among high school students, this was endorsed by 52 percent of grade ten students, 51 percent of grade eleven students, and 51 percent of grade twelve students.

For the second survey question, “Were any questions on the test different from the types of questions you see in science classes?,” 48 percent to 55 percent of students across grade levels agreed that some of the questions on the test were different. Across grade levels, 31 percent to 39 percent of students responded that most of the questions on the test were different. These percentages were higher for grades five and eight than for the high school grade levels when responding that some of the questions on the test were different, and lower for grades five and eight for the response that most of the questions were different.

The third survey question, “How hard were questions on this test compared to questions you see in science classes?,” resulted in 47 percent to 60 percent of students across grade levels felt that the test questions were about as hard as the questions in their science classes. Across grade levels, 27 percent to 42 percent of students felt that the test questions were harder than most questions in their science classes. The percentages in grades five and eight were higher when responding that the questions were about as hard, and the percentages for the high school grade band were higher when responding that the questions were harder.

For the fourth survey question, “Do you think you will be enrolling in any more science classes in high school?,” 66 percent of grade ten students and 49 percent of grade level eleven students responded that they would enroll in more science classes during high school.

The correlations between student survey responses and their scale scores across grade levels ranged from -0.33 to 0.28, indicating there was no clear relationship between student responses on the survey and their performance on the CAST.

### Appendix 9.A: Results of Student Survey

**Note:** “N/A” is not applicable and “SD” is standard deviation.

Table 9.A. Scale Score Distribution by Responses to Question 1 for Grade Five

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Did you learn about the topics on the test in your science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| My classes taught me about ALL of the topics on the test. | 11,865 | 19.5 | 196 | 22.8 | N/A |
| My classes taught me about MOST of the topics on the test, but not all of them. | 28,298 | 46.5 | 203 | 22.2 | N/A |
| My classes taught me about SOME of the topics on the test, but not most of them. | 16,342 | 26.9 | 199 | 21.1 | N/A |
| My classes did NOT teach me most of the topics on the test. | 4,330 | 7.1 | 197 | 21.1 | N/A |
| **Total** | **60,835** | **100** | **200** | **22.1** | **-0.01** |

\* R is the correlation between survey responses (1 = NOT, 2 = SOME, 3 = MOST, and 4 = ALL) and the CAST scale scores.

Table 9.A. Scale Score Distribution by Responses to Question 2 for Grade Five

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Were any questions on the test different from the types of questions you see in science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| ALL of the questions on the test were different. | 7,696 | 12.7 | 188 | 20.7 | N/A |
| MOST of the questions on the test were different. | 18,739 | 30.8 | 199 | 22.0 | N/A |
| SOME of the questions on the test were different. | 31,690 | 52.1 | 204 | 21.4 | N/A |
| NONE of the questions on the test were different. | 2,708 | 4.5 | 197 | 23.2 | N/A |
| **Total** | **60,833** | **100** | **200** | **22.1** | **0.19** |

\* R is the correlation between survey responses (1 = ALL, 2 = MOST, 3 = SOME, and 4 = NONE) and the CAST scale scores.

Table 9.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 science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| They were HARDER than most questions in my science classes. | 16,194 | 26.6 | 196 | 22.0 | N/A |
| They were ABOUT AS HARD as the questions in my science classes. | 36,502 | 60.0 | 202 | 21.8 | N/A |
| They were EASIER than most questions in my science classes. | 8,136 | 13.4 | 198 | 22.5 | N/A |
| **Total** | **60,832** | **100** | **200** | **22.1** | **0.06** |

\* R is the correlation between survey responses (1 = HARDER, 2 = ABOUT AS HARD, and 3 = EASIER) and the CAST scale scores.

Table 9.A. Scale Score Distribution by Responses to Question 1 for Grade Eight

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Did you learn about the topics on the test in your science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| My classes taught me about ALL of the topics on the test. | 10,758 | 15.5 | 400 | 23.5 | N/A |
| My classes taught me about MOST of the topics on the test, but not all of them. | 35,824 | 51.6 | 404 | 22.6 | N/A |
| My classes taught me about SOME of the topics on the test, but not most of them. | 19,600 | 28.2 | 397 | 20.1 | N/A |
| My classes did NOT teach me most of the topics on the test. | 3,257 | 4.7 | 391 | 19.5 | N/A |
| **Total** | **69,439** | **100** | **401** | **22.2** | **0.11** |

\* R is the correlation between survey responses (1 = NOT, 2 = SOME, 3 = MOST, and 4 = ALL) and CAST scale scores.

Table 9.A. Scale Score Distribution by Responses to Question 2 for Grade Eight

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Were any questions on the test different from the types of questions you see in science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| ALL of the questions on the test were different. | 5,881 | 8.5 | 387 | 19.9 | N/A |
| MOST of the questions on the test were different. | 22,745 | 32.8 | 399 | 21.4 | N/A |
| SOME of the questions on the test were different. | 37,974 | 54.7 | 404 | 21.9 | N/A |
| NONE of the questions on the test were different. | 2,831 | 4.1 | 400 | 25.1 | N/A |
| **Total** | **69,431** | **100** | **401** | **22.2** | **0.18** |

\* R is the correlation between survey responses (1 = ALL, 2 = MOST, 3 = SOME, and 4 = NONE) and the CAST scale scores.

Table 9.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 science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| They were HARDER than most questions in my science classes. | 26,603 | 38.3 | 397 | 21.0 | N/A |
| They were ABOUT AS HARD as the questions in my science classes. | 37,637 | 54.2 | 403 | 22.3 | N/A |
| They were EASIER than most questions in my science classes. | 5,188 | 7.5 | 399 | 25.2 | N/A |
| **Total** | **69,428** | **100** | **401** | **22.2** | **0.09** |

\* R is the correlation between survey responses (1 = HARDER, 2 = ABOUT AS HARD, and 3 = EASIER) and the CAST scale scores.

Table 9.A. Scale Score Distribution by Responses to Question 1 for Grade Ten

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Did you learn about the topics on the test in your science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| My classes taught me about ALL of the topics on the test. | 154 | 7.1 | 600 | 25.9 | N/A |
| My classes taught me about MOST of the topics on the test, but not all of them. | 971 | 44.7 | 611 | 22.5 | N/A |
| My classes taught me about SOME of the topics on the test, but not most of them. | 852 | 39.2 | 603 | 20.2 | N/A |
| My classes did NOT teach me most of the topics on the test. | 194 | 8.9 | 594 | 19.5 | N/A |
| **Total** | **2,171** | **100** | **606** | **22.4** | **0.17** |

\* R is the correlation between survey responses (1 = NOT, 2 = SOME, 3 = MOST, and 4 = ALL) and CAST scale scores.

Table 9.A. Scale Score Distribution by Responses to Question 2 for Grade Ten

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Were any questions on the test different from the types of questions you see in science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| ALL of the questions on the test were different. | 215 | 9.9 | 589 | 19.7 | N/A |
| MOST of the questions on the test were different. | 846 | 39.0 | 606 | 21.4 | N/A |
| SOME of the questions on the test were different. | 1,034 | 47.7 | 609 | 21.8 | N/A |
| NONE of the questions on the test were different. | 74 | 3.4 | 598 | 24.4 | N/A |
| **Total** | **2,169** | **100** | **606** | **22.4** | **0.18** |

\* R is the correlation between survey responses (1 = ALL, 2 = MOST, 3 = SOME, and 4 = NONE) and the CAST scale scores.

Table 9.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 science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| They were HARDER than most questions in my science classes. | 904 | 41.7 | 599 | 20.4 | N/A |
| They were ABOUT AS HARD as the questions in my science classes. | 1,025 | 47.3 | 608 | 21.8 | N/A |
| They were EASIER than most questions in my science classes. | 240 | 11.1 | 618 | 24.4 | N/A |
| **Total** | **2,169** | **100** | **606** | **22.4** | **0.28** |

\* R is the correlation between survey responses (1 = HARDER, 2 = ABOUT AS HARD, and 3 = EASIER) and the CAST scale scores.

Table 9.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 | R\* |
| Yes | 1,428 | 65.8 | 611 | 22.3 | N/A |
| No | 741 | 34.2 | 595 | 18.5 | N/A |
| **Total** | **2,169** | **100** | **606** | **22.4** | **-0.33** |

\* R is the correlation between survey responses (1 = No and 2 = Yes) and the CAST scale scores.

Table 9.A. Scale Score Distribution by Responses to Question 1 for Grade Eleven

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Did you learn about the topics on the test in your science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| My classes taught me about ALL of the topics on the test. | 3,921 | 9.4 | 605 | 25.0 | N/A |
| My classes taught me about MOST of the topics on the test, but not all of them. | 17,464 | 41.8 | 608 | 22.5 | N/A |
| My classes taught me about SOME of the topics on the test, but not most of them. | 16,114 | 38.5 | 600 | 19.6 | N/A |
| My classes did NOT teach me most of the topics on the test. | 4,308 | 10.3 | 591 | 17.9 | N/A |
| **Total** | **41,807** | **100** | **603** | **21.9** | **0.21** |

\* R is the correlation between survey responses (1 = NOT, 2 = SOME, 3 = MOST, and 4 = ALL) and CAST scale scores.

Table 9.A. Scale Score Distribution by Responses to Question 2 for Grade Eleven

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Were any questions on the test different from the types of questions you see in science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| ALL of the questions on the test were different. | 4,746 | 11.4 | 589 | 18.8 | N/A |
| MOST of the questions on the test were different. | 15,356 | 36.7 | 602 | 20.6 | N/A |
| SOME of the questions on the test were different. | 20,031 | 47.9 | 607 | 21.6 | N/A |
| NONE of the questions on the test were different. | 1,668 | 4.0 | 603 | 25.3 | N/A |
| **Total** | **41,801** | **100** | **603** | **21.9** | **0.23** |

\* R is the correlation between survey responses (1 = ALL, 2 = MOST, 3 = SOME, and 4 = NONE) and the CAST scale scores.

Table 9.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 science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| They were HARDER than most questions in my science classes. | 17,559 | 42.0 | 597 | 19.5 | N/A |
| They were ABOUT AS HARD as the questions in my science classes. | 20,215 | 48.4 | 606 | 21.5 | N/A |
| They were EASIER than most questions in my science classes. | 4,024 | 9.6 | 614 | 25.7 | N/A |
| **Total** | **41,798** | **100** | **603** | **21.9** | **0.25** |

\* R is the correlation between survey responses (1 = HARDER, 2 = ABOUT AS HARD, and 3 = EASIER) and the CAST scale scores.

Table 9.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 | R\* |
| Yes | 20,407 | 48.8 | 608 | 22.3 | N/A |
| No | 21,387 | 51.2 | 598 | 20.4 | N/A |
| **Total** | **41,794** | **100** | **603** | **21.9** | **-0.22** |

\* R is the correlation between survey responses (1 = No and 2 = Yes) and the CAST scale scores.

Table 9.A. Scale Score Distribution by Responses to Question 1 for Grade Twelve

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Did you learn about the topics on the test in your science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| My classes taught me about ALL of the topics on the test. | 6,188 | 10.5 | 603 | 24.2 | N/A |
| My classes taught me about MOST of the topics on the test, but not all of them. | 23,860 | 40.4 | 606 | 22.0 | N/A |
| My classes taught me about SOME of the topics on the test, but not most of them. | 22,224 | 37.6 | 598 | 18.6 | N/A |
| My classes did NOT teach me most of the topics on the test. | 6,844 | 11.6 | 589 | 16.6 | N/A |
| **Total** | **59,116** | **100** | **601** | **21.2** | **0.22** |

\* R is the correlation between survey responses (1 = NOT, 2 = SOME, 3 = MOST, and 4 = ALL) and CAST scale scores.

Table 9.A. Scale Score Distribution by Responses to Question 2 for Grade Twelve

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Were any questions on the test different from the types of questions you see in science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| ALL of the questions on the test were different. | 7,108 | 12.0 | 587 | 17.1 | N/A |
| MOST of the questions on the test were different. | 20,864 | 35.3 | 599 | 19.8 | N/A |
| SOME of the questions on the test were different. | 28,211 | 47.7 | 606 | 21.0 | N/A |
| NONE of the questions on the test were different. | 2,932 | 5.0 | 603 | 24.8 | N/A |
| **Total** | **59,115** | **100** | **601** | **21.2** | **0.25** |

\* R is the correlation between survey responses (1 = ALL, 2 = MOST, 3 = SOME, and 4 = NONE) and the CAST scale scores.

Table 9.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 science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| They were HARDER than most questions in my science classes. | 23,439 | 39.7 | 595 | 18.4 | N/A |
| They were ABOUT AS HARD as the questions in my science classes. | 29,760 | 50.3 | 604 | 20.7 | N/A |
| They were EASIER than most questions in my science classes. | 5,915 | 10.0 | 612 | 25.8 | N/A |
| **Total** | **59,114** | **100** | **601** | **21.2** | **0.27** |

\*R is the correlation between survey responses (1 = HARDER, 2 = ABOUT AS HARD, and 3 = EASIER) and the CAST scale scores.

Table 9.A. Scale Score Distribution by Responses to Question 1 for High School

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Did you learn about the topics on the test in your science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| My classes taught me about ALL of the topics on the test. | 10,263 | 10.0 | 604 | 24.5 | N/A |
| My classes taught me about MOST of the topics on the test, but not all of them. | 42,295 | 41.0 | 607 | 22.3 | N/A |
| My classes taught me about SOME of the topics on the test, but not most of them. | 39,190 | 38.0 | 599 | 19.1 | N/A |
| My classes did NOT teach me most of the topics on the test. | 11,346 | 11.0 | 590 | 17.2 | N/A |
| **Total** | **103,094** | **100** | **602** | **21.5** | **0.21** |

\* R is the correlation between survey responses (1 = NOT, 2 = SOME, 3 = MOST, and 4 = ALL) and CAST scale scores.

Table 9.A. Scale Score Distribution by Responses to Question 2 for High School

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Were any questions on the test different from the types of questions you see in science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| ALL of the questions on the test were different. | 12,069 | 11.7 | 588 | 17.9 | N/A |
| MOST of the questions on the test were different. | 37,066 | 36.0 | 600 | 20.2 | N/A |
| SOME of the questions on the test were different. | 49,276 | 47.8 | 606 | 21.3 | N/A |
| NONE of the questions on the test were different. | 4,674 | 4.5 | 603 | 25.0 | N/A |
| **Total** | **103,085** | **100** | **602** | **21.5** | **0.24** |

\* R is the correlation between survey responses (1 = ALL, 2 = MOST, 3 = SOME, and 4 = NONE) and the CAST scale scores.

Table 9.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 science classes? | N | Percent | Scale Score Mean | Scale Score SD | R\* |
| They were HARDER than most questions in my science classes. | 41,902 | 40.6 | 596 | 19.0 | N/A |
| They were ABOUT AS HARD as the questions in my science classes. | 51,000 | 49.5 | 605 | 21.1 | N/A |
| They were EASIER than most questions in my science classes. | 10,179 | 9.9 | 613 | 25.8 | N/A |
| **Total** | **103,081** | **100** | **602** | **21.5** | **0.26** |

\* R is the correlation between survey responses (1 = HARDER, 2 = ABOUT AS HARD, and 3 = EASIER) and the CAST scale scores.

Table 9.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 | R\* |
| Yes | 21,835 | 49.7 | 608 | 22.3 | N/A |
| No | 22,128 | 50.3 | 598 | 20.3 | N/A |
| **Total** | **43,963** | **100** | **603** | **21.9** | **-0.23** |

\* R is the correlation between survey responses (1 = No and 2 = Yes) and the CAST scale scores.

## Continuous and Systematic Improvement

This chapter documents the process whereby continuous improvements are ensured and the results of this process in the current year.

### Test Design

ETS works in collaboration with the California Department of Education (CDE) in planning, proposing, evaluating, and improving the California Science Test (CAST) design.

The operational test forms for the 2020–2021 CAST administration were delivered in accordance with the approved blueprint, as were the test forms for previous administrations. As a result of related research studies, ETS, in collaboration with the CDE, developed an updated blueprint that was presented to the California State Board of Education (SBE) in January 2020. The updated blueprint was approved and was intended to be implemented beginning with the 2020–2021 administration. However, because of the suspension of the 2019–2020 administration, the implementation of the updated blueprint was postponed to the 2021–2022 administration. The updated blueprint differs from the current blueprint in a few key areas, such as that Segment B will contain three performance tasks (PTs).

### Item Development

For the 2020–2021 item development cycle, the ETS content team used item content specifications 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 shared with the public during the 2019–2020 administration and met with positive feedback.

Item performance data from the 2017–2018 CAST field test administration provided additional information with which to review items during and after a data review meeting with teachers in the field in 2018. This additional information was key to making constructive changes to item development processes internal to ETS.

Field test item data showed what worked and what did not work, especially in terms of the language used in items. Both positive and negative exemplar items from the 2018–‍2019 operational administration 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.

Another process improvement was to field-test parallel PTs 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 SBE-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.

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 2021–2022 administration.

Table 10.1 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 the 2018–2019 cycle to the 2020–2021 cycle, the low rejection rates for both grade levels and the high school grade band remained nearly the same. (As noted in prior sections, the suspension of the 2019–2020 administration resulted in very limited psychometric analysis.)

Table 10.1 Item Development Results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Total Number of Unique Items Field-Tested in 2018–2019 Administration | Number of Rejected Items From 2018–2019 | Rejection Rate From 2019 Data Review | Total Number of Unique Items Field-Tested in the 2020–2021 Administration | Number of Rejected Items From 2020–2021 Cycle | Rejection Rate From 2021 Data Review |
| Grade 5 | 234 | 7 | 3% | 156 | 5 | 3% |
| Grade 8 | 228 | 11 | 5% | 145 | 8 | 6% |
| High school | 224 | 7 | 3% | 184 | 4 | 2% |

### Administration and Test Delivery

ETS continues to improve in the area of test delivery, and the 2019–2020 administration introduced an improved progress bar for the student as the student progressed through the test. Due to the suspension of testing in 2019–2020, the improved progress bar was not used by CAST test-takers until the 2020–2021 administration.

### Constructed-Response Item Scoring

#### Human Scoring Activities

Continuous improvements for the 2020–2021 administration included that rater agreement and validity statistics for rater agreement were monitored during each scoring shift. Scoring leaders provided feedback to ETS Assessment and Learning Technology Research & Development to determine what, if any, adjustments to training or samples were to be made.

ETS will use standardized training to assist the scoring leader in using the performance indicator panels, which allow easier access to quantitative feedback regarding individual raters.

Improved training for scoring leaders will be conducted via online learning courses, as opposed to the WebEx sessions used previously. Online learning courses provide the following expected benefits:

* Standardized explanation on what information is available within the performance indicator panel as well as how to use the information
* Standardized format for providing feedback to individual raters to ensure the area of improvement needed is clear and consistent regardless of which score leader may be monitoring an individual rater on any given day
* Automatic restriction of scoring leaders from monitoring raters until training requirements from 2020–2021 have been satisfied (previously done manually)

### Psychometric Analyses

Preequating was implemented in the 2020–2021 test administration to allow an early reporting timeline. For details about the benefits, feasibility, and concerns of the preequating plan, refer to the *Feasibility of a Preequating Plan for the California Science Test (CAST) for the 2019–2020 Administration* (ETS, 2020).

A comprehensive scoring quality control process, necessary for the preequating process, has been developed and was implemented for the 2020–2021 test administration to ensure the accuracy of all scores in early reporting.

The 2020–2021 administration was unique, given the ongoing impact of the novel coronavirus disease 2019 pandemic. This resulted in the CAST being administered in person and remotely, depending on testing needs within the local educational agencies (LEAs). A study was conducted to evaluate the comparability of student scores between the two testing environments. Refer to subsection[*5.1.1 Remote and In-Person Testing*](#_Remote_and_In-Person_1) and section [*7.8 Score Comparability—Remote Versus In-Person Testing Analyses*](#_Score_Comparability—Remote_Versus) for a summary of this study.

ETS will explore the possibility of providing more graphs and plots, rather than primarily using tables, in future technical reports, to make the technical reports more visualized and user friendly.

### Accessibility

ETS increased the number of accessibility resources available, as evidenced by the expansion of the multiplication table to a 12 × 12 table, updating braille embossing to prefetch more items to be embossed, and permitting students using the scratch paper universal tool to create graphic organizers.

As for the items themselves, ETS remains focused on making items accessible from the outset, reducing the need to provide extensive adaptations to make the items accessible to students with visual impairment.

### Feedback for Continuous Improvement Survey

ETS annually solicits feedback from educators through a survey that allows the CDE and ETS to focus on continuous improvement. LEA and test site staff, as well as test administrators and test examiners, were invited to participate in the 2020–‍21 Feedback for Continuous Improvement Survey. Because of the unique nature of the 2020–2021 administration year and the option to administer assessments remotely or administer local assessments for Smarter Balanced for English language arts/literacy and mathematics (the CAST did not offer a local assessment), the survey centered on preparation, training, and test administration, including remote testing. More than 1,500 California educators provided specific, actionable insights about their testing experience. The CDE and ETS use key recommendations from educators to implement positive changes in the following administration year.

This survey gathered information and data from educators who were part of the administration of the California Assessment of Student Performance and Progress (CAASPP) and the English Language Proficiency Assessments for California (ELPAC). Its goal was to highlight successes and identify areas for improvement.

More than half (54%) of survey respondents used both remote and in-person options to complete testing. Overall, California educators continued to express positive experiences in their preparations for CAASPP and the ELPAC. Although the 2020–2021 administration included the daunting task of remote testing, educators felt that the resources and training materials they were given were useful in preparing them and their students for test administration. Their feedback generally described smooth preparation, training, support, and assessment administration experiences. Also, educators provided valuable feedback for potential improvements for future administrations based upon lessons learned.

The majority of respondents (64%) felt prepared for administering remote testing. On average, LEA CAASPP and ELPAC coordinators reported feeling prepared at a higher rate than those in site-level roles, such as test administrators and test examiners. Only a small percentage of respondents (6%) reported not feeling prepared. In regard to remote testing, educators felt they could benefit from more troubleshooting resources to deal with the technical difficulties that arise during remote testing. They indicated that a simplified process for logging on to the secure browser would be helpful for students, particularly English learner students. When asked about training preferences, respondents indicated that self-paced online trainings were preferable over in-person workshops and live virtual trainings. For live virtual trainings, Zoom is the preferred platform.

Refer to the 2020–21 CAASPP and ELPAC Feedback for Continuous Improvement Survey and Focus Groups Report for detailed information (CDE, 2021).

### References

Educational Testing Service. (2020). *Feasibility of a preequating plan for the California Science Test (CAST) for the 2019–2020 administration*. [Unpublished memorandum]. Princeton, NJ: Educational Testing Service.

California Department of Education. (2021). *2020–21 CAASPP and ELPAC feedback for continuous improvement survey and focus groups report*. Sacramento, CA: California Department of Education.

1. This information 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. This technical report is based on the version of the *Usability, Accessibility, and Accommodations Guidelines* that was available during the 2020–2021 CAASPP administration. [↑](#footnote-ref-4)
4. During the early years of CAST administration, meeting the specifications was more challenging as the item bank was being developed. [↑](#footnote-ref-5)
5. Also refer to subsection [*5.1.1 Remote and In-Person Testing*](#_Remote_and_In-Person_1) for additional information about testing locations. [↑](#footnote-ref-6)
6. These were the same as the calculators used during administration of the Smarter Balanced for Mathematics Summative Assessment. [↑](#footnote-ref-7)
7. The expandable items universal tool was turned on by the test administrator in the Test Administrator Interface. [↑](#footnote-ref-8)
8. These were the same as the mathematics tools used during administration of the Smarter Balanced for Mathematics Summative Assessment. [↑](#footnote-ref-9)
9. PDFs of the science charts are available for download from the California Science Test web page on the CAASPP website. [↑](#footnote-ref-10)
10. The *Crosswalk* has since been replaced with the Accessibility Strategies web page on the Tools for Teachers website. [↑](#footnote-ref-11)
11. 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 (Vapnick, 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, Burgess, et al., 1996). [↑](#footnote-ref-12)
12. The timing data is based on capturing the amount of time spent on answering the item(s) on each page. [↑](#footnote-ref-13)
13. Testing time analyses combined data from students who tested in person and remotely. Refer to the *Results from the Analysis of Remote Testing of the 2020–2021 California Science Tests* (CDE, 2022) for detailed information on testing time for these cohorts. [↑](#footnote-ref-14)
14. Prior Smarter Balanced mathematics scores were not used for high school, because nearly all students in grades eleven and twelve were not administered a Smarter Balanced Summative Assessment in 2018–2019, and most grade twelve students had not been administered a Smarter Balanced assessment since the 2016–2017 administration. [↑](#footnote-ref-15)