

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



# California Assessment of Student Performance and Progress California Spanish Assessment 2021–‍22 Technical Report

**Submitted July 13, 2023**

**By ETS**



**Contract #CN150012**

Table of Contents

[Chapter 1: Introduction 1](#_Toc138337683)

[1.1. Background 1](#_Toc138337684)

[1.2. Test Purpose(s) 2](#_Toc138337685)

[1.3. Test Design 2](#_Toc138337686)

[1.4. Intended Population 2](#_Toc138337687)

[1.5. Intended Use and Purpose of Test Scores 3](#_Toc138337688)

[1.6. Testing Window 3](#_Toc138337689)

[1.7. Impact of the Novel Coronavirus Disease 2019 Pandemic 4](#_Toc138337690)

[1.7.1. Remote Testing Flexibility 4](#_Toc138337691)

[1.8. Significant CSA Developments in 2021–22 4](#_Toc138337692)

[1.8.1. Accessibility Resources 4](#_Toc138337693)

[1.8.2. Reused Forms for High School in 2021‒22 4](#_Toc138337694)

[1.8.3. High School Postequating Procedures 5](#_Toc138337695)

[1.9. Groups and Organizations Involved with the CAASPP System 5](#_Toc138337696)

[1.9.1. California State Board of Education 5](#_Toc138337697)

[1.9.2. California Department of Education 5](#_Toc138337698)

[1.9.3. California Educators 6](#_Toc138337699)

[1.9.4. Contractors 6](#_Toc138337700)

[1.10. Systems Overview and Functionality 7](#_Toc138337701)

[1.10.1. Test Operations Management System 7](#_Toc138337702)

[1.10.2. Test Delivery System 8](#_Toc138337703)

[1.10.3. Practice and Training Tests 8](#_Toc138337704)

[1.10.4. California Educator Reporting System 8](#_Toc138337705)

[1.10.5. Test Results for California’s Assessments Website 9](#_Toc138337706)

[1.11. Overview of the Technical Report 9](#_Toc138337707)

[Reference 11](#_Toc138337708)

[Chapter 2: Overview of CSA Processes 12](#_Toc138337709)

[2.1. Item Development 12](#_Toc138337710)

[2.1.1. Item Format 12](#_Toc138337711)

[2.1.2. Item Specifications 12](#_Toc138337712)

[2.2. Test Assembly 13](#_Toc138337713)

[2.2.1. Test Length 13](#_Toc138337714)

[2.2.2. Test Blueprints 14](#_Toc138337715)

[2.2.3. Item Selection 14](#_Toc138337716)

[2.3. Test Administration 14](#_Toc138337717)

[2.3.1. Test Security and Confidentiality 14](#_Toc138337718)

[2.3.2. Procedures to Maintain Standardization 15](#_Toc138337719)

[2.4. Fairness and Accessibility 15](#_Toc138337720)

[2.4.1. Overview 15](#_Toc138337721)

[2.4.2. Universal Tools, Designated Supports, and Accommodations 16](#_Toc138337722)

[2.4.3. Description of Differential Item Functioning 16](#_Toc138337723)

[2.5. Scores 16](#_Toc138337724)

[2.5.1. Score Reporting 17](#_Toc138337725)

[2.5.2. Aggregation Procedures 17](#_Toc138337726)

[2.6. Psychometric Analyses 17](#_Toc138337727)

[2.6.1. Description of Classical Item Analyses 17](#_Toc138337728)

[2.6.2. Description of Item Response Theory Analyses 18](#_Toc138337729)

[References 19](#_Toc138337730)

[Chapter 3: Item Development and Review 20](#_Toc138337731)

[3.1. Guidelines 20](#_Toc138337732)

[3.1.1. Overview 20](#_Toc138337733)

[3.1.2. Plan 21](#_Toc138337734)

[3.1.3. Process 21](#_Toc138337735)

[3.1.4. Item Specifications 22](#_Toc138337736)

[3.1.5. Item Format 23](#_Toc138337737)

[3.1.6. Item Types 23](#_Toc138337738)

[3.1.7. Item Banking 25](#_Toc138337739)

[3.1.8. Recruitment and Selection of Item Writers 25](#_Toc138337740)

[3.1.9. Item Writer Training 26](#_Toc138337741)

[3.2. ETS Item Review Process 26](#_Toc138337742)

[3.2.1. Overview 26](#_Toc138337743)

[3.2.2. ETS Content Review 27](#_Toc138337744)

[3.2.3. ETS Accessibility Review 27](#_Toc138337745)

[3.2.4. ETS Editorial Review 28](#_Toc138337746)

[3.2.5. ETS Sensitivity and Fairness Review 28](#_Toc138337747)

[3.3. California Department of Education Review 28](#_Toc138337748)

[3.4. California Educator Review 28](#_Toc138337749)

[3.4.1. Composition of Item Review Panels 29](#_Toc138337750)

[3.4.2. Meetings for Review of CSA Field Test Items 31](#_Toc138337751)

[3.5. Data Review Meeting 31](#_Toc138337752)

[References 33](#_Toc138337753)

[Chapter 4: Test Assembly 34](#_Toc138337754)

[4.1. Test Design 34](#_Toc138337755)

[4.2. Test Blueprints and Other Content Specifications 34](#_Toc138337756)

[4.2.1. Test Blueprints 34](#_Toc138337757)

[4.3. Test Production Process 35](#_Toc138337758)

[4.3.1. Content Criteria 35](#_Toc138337759)

[4.3.2. Test Forms 35](#_Toc138337760)

[4.3.3. Psychometric Criteria and Identification of Eligible Items 35](#_Toc138337761)

[4.3.4. Content Review of Forms 37](#_Toc138337762)

[4.3.5. California Department of Education Forms Review 37](#_Toc138337763)

[4.3.6. Configuration of the Test Delivery System 38](#_Toc138337764)

[4.3.7. Delivery of General Forms 38](#_Toc138337765)

[4.4. Special Version Forms 38](#_Toc138337766)

[4.4.1. Braille Form 38](#_Toc138337767)

[4.4.2. Accommodated Forms Other Than Braille 39](#_Toc138337768)

[References 40](#_Toc138337769)

[Appendix 4.A: Test Blueprints 41](#_Toc138337770)

[Chapter 5: Test Administration 45](#_Toc138337771)

[5.1. Overview 45](#_Toc138337772)

[5.1.1. In-Person and Remote Testing 45](#_Toc138337773)

[5.2. User Roles and Standardization 45](#_Toc138337774)

[5.2.1. Local Educational Agency CAASPP Coordinator 45](#_Toc138337775)

[5.2.2. CAASPP Test Site Coordinator 46](#_Toc138337776)

[5.2.3. Test Administrators 47](#_Toc138337777)

[5.2.4. Instructions for Test Administrators 47](#_Toc138337778)

[5.3. Local Educational Agency Training 49](#_Toc138337779)

[5.3.1. Synchronous and Asynchronous Training 49](#_Toc138337780)

[5.3.2. Videos and Guides 50](#_Toc138337781)

[5.3.3. Training for Proper Identification and Assignment of Designated Supports and Accommodations 50](#_Toc138337782)

[5.3.4. Feedback for Continuous Improvement Survey 51](#_Toc138337783)

[5.4. Accessibility Resources 52](#_Toc138337784)

[5.4.1. Accessibility Resource Categories 52](#_Toc138337785)

[5.4.2. Identification and Selection 55](#_Toc138337786)

[5.4.3. Assignment 55](#_Toc138337787)

[5.4.4. Delivery of Embedded and Non-Embedded Resources to Students 56](#_Toc138337788)

[5.4.5. Usage of Designated Supports and Accommodations 56](#_Toc138337789)

[5.5. Practice and Training Tests 61](#_Toc138337790)

[5.6. Test Security and Confidentiality 61](#_Toc138337791)

[5.6.1. ETS’ Office of Testing Integrity 62](#_Toc138337792)

[5.6.2. Procedures to Maintain Standardization of Test Security 62](#_Toc138337793)

[5.6.3. Test Security Monitoring 63](#_Toc138337794)

[5.6.4. Security of Electronic Files Using a Firewall 63](#_Toc138337795)

[5.6.5. Transfer of Scores via Secure Data Exchange 64](#_Toc138337796)

[5.6.6. Data Management in the Secure Database 64](#_Toc138337797)

[5.6.7. Statistical Analysis on Secure Servers 64](#_Toc138337798)

[5.6.8. Student Confidentiality 65](#_Toc138337799)

[5.6.9. Student Test Results 65](#_Toc138337800)

[5.6.10. Security and Test Administration Incident Reporting System Process 66](#_Toc138337801)

[5.6.11. Appeals 68](#_Toc138337802)

[References 71](#_Toc138337803)

[Appendix 5.A: Accessibility Resource Assignment 72](#_Toc138337804)

[Chapter 6: Standard Setting 78](#_Toc138337805)

[6.1. Description 78](#_Toc138337806)

[Reference 79](#_Toc138337807)

[Chapter 7: Scoring and Reporting 80](#_Toc138337808)

[7.1. Student Test Scores 80](#_Toc138337809)

[7.1.1. Theta Scores 80](#_Toc138337810)

[7.1.2. Scale Scores 81](#_Toc138337811)

[7.1.3. Score Reporting Ranges 82](#_Toc138337812)

[7.1.4. Scoring of Incomplete Cases 82](#_Toc138337813)

[7.2. Overview of Score Aggregation Procedures 83](#_Toc138337814)

[7.2.1. Individual Student Score Distributions and Summary Statistics 84](#_Toc138337815)

[7.2.2. Demographic Student Group Summaries 85](#_Toc138337816)

[7.3. Reports Produced and Scores for Each Report 86](#_Toc138337817)

[7.3.1. Online Reporting 86](#_Toc138337818)

[7.3.2. Special Cases 87](#_Toc138337819)

[7.3.3. Types of Score Reports 87](#_Toc138337820)

[7.3.4. Score Report Applications 89](#_Toc138337821)

[7.3.5. Criteria for Interpreting Test Scores 89](#_Toc138337822)

[7.3.6. Criteria for Interpreting Score Reports 90](#_Toc138337823)

[References 91](#_Toc138337824)

[Appendix 7.A: Raw Score, Theta Score, Scale Score, and Conditional Standard Error of Measurement Distributions of Students Taking Each Test 92](#_Toc138337825)

[Appendix 7.B: Demographic Student Group Summaries of Participation 107](#_Toc138337826)

[Appendix 7.C: Demographic Student Group Summaries of Performance 121](#_Toc138337827)

[Chapter 8: Psychometric Analyses 136](#_Toc138337828)

[8.1. Overview 136](#_Toc138337829)

[8.1.1. Summary of the Analyses 136](#_Toc138337830)

[8.1.2. Samples Used for Analyses 136](#_Toc138337831)

[8.1.3. Test-Taking Rates 137](#_Toc138337832)

[8.2. Classical Item Analyses 138](#_Toc138337833)

[8.2.1. Classical Item Difficulty Indices (*p*-value and Average Item Score) 138](#_Toc138337834)

[8.2.2. Item-Total Correlation 139](#_Toc138337835)

[8.2.3. Distribution of Item Scores 140](#_Toc138337836)

[8.2.4. Omit Rates 140](#_Toc138337837)

[8.2.5. Completion Rates 141](#_Toc138337838)

[8.2.6. Distractor Analyses 141](#_Toc138337839)

[8.2.7. Summary of Classical Item Analyses Flagging Criteria 141](#_Toc138337840)

[8.2.8. Classical Item Analyses Results 142](#_Toc138337841)

[8.3. Differential Item Functioning Analyses 145](#_Toc138337842)

[8.3.1. Differential Item Functioning Procedure for Dichotomous Items 146](#_Toc138337843)

[8.3.2. Differential Item Functioning Procedure for Polytomous Items 146](#_Toc138337844)

[8.3.3. Classification 147](#_Toc138337845)

[8.3.4. Differential Item Functioning Analysis Results 148](#_Toc138337846)

[8.4. Item Response Theory Analyses 150](#_Toc138337847)

[8.4.1. Item Response Theory Model 150](#_Toc138337848)

[8.4.2. Data Preparation 151](#_Toc138337849)

[8.4.3. Equating 151](#_Toc138337850)

[8.4.4. Calibration and Linking for the Field Test Items 156](#_Toc138337851)

[8.4.5. Parameter Estimates 158](#_Toc138337852)

[8.4.6. Scaling the Scores 159](#_Toc138337853)

[8.5. Testing Time Analyses 162](#_Toc138337854)

[8.6. Reliability Analyses 162](#_Toc138337855)

[8.6.1. Sample for Reliability Analyses 163](#_Toc138337856)

[8.6.2. Reliability Measures 163](#_Toc138337857)

[8.6.3. Standard Error of Measurement 164](#_Toc138337858)

[8.6.4. Student Group Reliabilities and Standard Errors of Measurement 164](#_Toc138337859)

[8.6.5. Conditional Standard Errors of Measurement 165](#_Toc138337860)

[8.6.6. Decision Classification Analyses 167](#_Toc138337861)

[8.7. Validity Evidence 168](#_Toc138337862)

[8.7.1. Evidence Based on the Design of the CSA 169](#_Toc138337863)

[8.7.2. Evidence Based on Content 171](#_Toc138337864)

[8.7.3. Evidence Based on the Response Processes 172](#_Toc138337865)

[8.7.4. Evidence Based on the Internal Structure 173](#_Toc138337866)

[References 175](#_Toc138337867)

[Accessibility Information 178](#_Toc138337868)

[Alternative Text for Equation 8.1 178](#_Toc138337869)

[Alternative Text for Equation 8.2 178](#_Toc138337870)

[Alternative Text for Equation 8.3 178](#_Toc138337871)

[Alternative Text for Equation 8.4 178](#_Toc138337872)

[Alternative Text for Equation 8.5 178](#_Toc138337873)

[Alternative Text for Equation 8.6 178](#_Toc138337874)

[Alternative Text for Equation 8.7 178](#_Toc138337875)

[Alternative Text for Equation 8.8 178](#_Toc138337876)

[Alternative Text for Equation 8.9 179](#_Toc138337877)

[Alternative Text for Equation 8.10 179](#_Toc138337878)

[Alternative Text for Equation 8.11 179](#_Toc138337879)

[Alternative Text for Equation 8.12 179](#_Toc138337880)

[Alternative Text for Equation 8.13 179](#_Toc138337881)

[Alternative Text for Equation 8.14 179](#_Toc138337882)

[Alternative Text for Equation 8.15 179](#_Toc138337883)

[Alternative Text for Equation 8.16 179](#_Toc138337884)

[Alternative Text for Equation 8.17 179](#_Toc138337885)

[Alternative Text for Equation 8.18 179](#_Toc138337886)

[Alternative Text for Equation 8.19 179](#_Toc138337887)

[Alternative Text for Equation 8.20 179](#_Toc138337888)

[Alternative Text for Equation 8.21 180](#_Toc138337889)

[Alternative Text for Equation 8.22 180](#_Toc138337890)

[Alternative Text for Equation 8.23 180](#_Toc138337891)

[Appendix 8.A: Classical Item Analyses 181](#_Toc138337892)

[Appendix 8.B: Item Response Theory Analyses 197](#_Toc138337893)

[Appendix 8.C: Testing Time Analyses 211](#_Toc138337894)

[Appendix 8.D: Reliability Analyses 215](#_Toc138337895)

[Appendix 8.E: Analyses of Classification 219](#_Toc138337896)

[Chapter 9: Quality Control Procedures 223](#_Toc138337897)

[9.1. Quality Control of Item Development 223](#_Toc138337898)

[9.2. Quality Control of Test Assembly and Delivery 223](#_Toc138337899)

[9.3. Quality Control of Test Materials 224](#_Toc138337900)

[9.3.1. Collecting Test Materials for Computer-based Assessments 224](#_Toc138337901)

[9.3.2. Processing Test Materials for Computer-based Assessments 224](#_Toc138337902)

[9.4. Quality Control of Test Administration 225](#_Toc138337903)

[9.5. Quality Control of Scoring 225](#_Toc138337904)

[9.5.1. Machine-Scoring Procedures 225](#_Toc138337905)

[9.5.2. Development of Scoring Specifications 226](#_Toc138337906)

[9.6. Quality Control of Psychometric Processes 226](#_Toc138337907)

[9.6.1. Scoring Verification 226](#_Toc138337908)

[9.6.2. Psychometric Analyses 226](#_Toc138337909)

[9.7. Quality Control of Reporting 227](#_Toc138337910)

[9.7.1. Exclusion of Student Scores from Summary Reports 228](#_Toc138337911)

[9.8. Quality Control of End-to-End Testing 228](#_Toc138337912)

[9.8.1. Computer-based Assessments 228](#_Toc138337913)

[Reference 230](#_Toc138337914)

[Chapter 10: Continuous and Systematic Improvement 231](#_Toc138337915)

[10.1. 2021–22 Feedback for Continuous Improvement Survey 231](#_Toc138337916)

[10.1.1. Outreach 231](#_Toc138337917)

[10.1.2. Communications 232](#_Toc138337918)

[10.2. Accessibility Resources 232](#_Toc138337919)

[Reference 233](#_Toc138337920)

List of Tables

[Acronyms and Initialisms Used in the *California Spanish Assessment Technical Report* xi](#_Toc138337921)

[Table 3.1 Number of Items Developed per Grade Level and Grade Band for the CSA 21](#_Toc138337922)

[Table 3.2 Item Types for the CSA 24](#_Toc138337923)

[Table 3.3 CSA Item Reviewer Qualifications 29](#_Toc138337924)

[Table 3.4 Status of Items After the 2020 IRM 30](#_Toc138337925)

[Table 3.5 Status of Items After the 2021 IRM 30](#_Toc138337926)

[Table 3.6 Item Data Review Results 32](#_Toc138337927)

[Table 4.1 Number of CSA Items to Administer per Form 34](#_Toc138337928)

[Table 4.2 Number of Forms and Items Reviewed Psychometrically 36](#_Toc138337929)

[Table 4.A.1 CSA Test Blueprint for Grade Span Three Through Five 42](#_Toc138337930)

[Table 4.A.2 CSA Test Blueprint for Grade Span Six Through Eight 43](#_Toc138337931)

[Table 4.A.3 CSA Test Blueprint for Grade Span Nine Through Twelve (High School) 44](#_Toc138337932)

[Table 5.2 Summary of Accommodations and Designated Supports Used by Students, by Grade Level or Grade Band 58](#_Toc138337933)

[Table 5.3 Types of Appeals 67](#_Toc138337934)

[Table 5.4 Number and Types of Incidents Submitted in STAIRS in the 2021–22 Test Administration 69](#_Toc138337935)

[Table 5.5 Number of Appeals Approved in STAIRS in the 2021–22 Test Administration 70](#_Toc138337936)

[Table 5.A.1 Special Services Summary, Grades Three Through Five 72](#_Toc138337937)

[Table 5.A.2 Special Services Summary, Grades Six Through Eight 74](#_Toc138337938)

[Table 5.A.3 Special Services Summary, High School 76](#_Toc138337939)

[Table 7.1 CSA Score Reporting Ranges by Grade Level or Grade Band 82](#_Toc138337940)

[Table 7.2 Rules for Incomplete Tests 83](#_Toc138337941)

[Table 7.3 Mean and SD of Theta Scores and Scale Scores 84](#_Toc138337942)

[Table 7.4 Numbers and Percentages of Students in Score Reporting Ranges 84](#_Toc138337943)

[Table 7.5 Demographic Student Groups to Be Reported 85](#_Toc138337944)

[Table 7.A.1 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for Grade Three 92](#_Toc138337945)

[Table 7.A.2 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for Grade Four 95](#_Toc138337946)

[Table 7.A.3 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for Grade Five 97](#_Toc138337947)

[Table 7.A.4 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for Grade Six 99](#_Toc138337948)

[Table 7.A.5 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for Grade Seven 101](#_Toc138337949)

[Table 7.A.6 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for Grade Eight 103](#_Toc138337950)

[Table 7.A.7 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for High School 105](#_Toc138337951)

[Table 7.B.1 Demographic Summary for Grade Three 107](#_Toc138337952)

[Table 7.B.2 Demographic Summary for Grade Four 109](#_Toc138337953)

[Table 7.B.3 Demographic Summary for Grade Five 111](#_Toc138337954)

[Table 7.B.4 Demographic Summary for Grade Six 113](#_Toc138337955)

[Table 7.B.5 Demographic Summary for Grade Seven 115](#_Toc138337956)

[Table 7.B.6 Demographic Summary for Grade Eight 117](#_Toc138337957)

[Table 7.B.7 Demographic Student Group Summary for High School 119](#_Toc138337958)

[Table 7.C.1 Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for Grade Three 121](#_Toc138337959)

[Table 7.C.2 Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for Grade Four 124](#_Toc138337960)

[Table 7.C.3 Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for Grade Five 126](#_Toc138337961)

[Table 7.C.4 Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for Grade Six 128](#_Toc138337962)

[Table 7.C.5 Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for Grade Seven 130](#_Toc138337963)

[Table 7.C.6 Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for Grade Eight 132](#_Toc138337964)

[Table 7.C.7 Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for High School, All Grade Levels 134](#_Toc138337965)

[Table 8.1 Sample Size of Data Files Used for Different Analyses 137](#_Toc138337966)

[Table 8.2 Total Number Tested in the 2021–22 Administration 138](#_Toc138337967)

[Table 8.3 Test-Taking Rates by California Region 138](#_Toc138337968)

[Table 8.4 Item Difficulty Distributions by Grade Level or Grade Band 142](#_Toc138337969)

[Table 8.5 Item Total Correlation Distributions by Grade Level or Grade Band 143](#_Toc138337970)

[Table 8.6 Flagged Items Summary in Each Form by Grade Level or Grade Band 144](#_Toc138337971)

[Table 8.7 Student Groups for DIF Comparison 145](#_Toc138337972)

[Table 8.8 DIF Categories for Dichotomous Items 147](#_Toc138337973)

[Table 8.9 DIF Categories for Polytomous Items 148](#_Toc138337974)

[Table 8.10 Gender DIF Classifications Summary by Grade Level or Grade Band 149](#_Toc138337975)

[Table 8.11 Final Linking Summary 154](#_Toc138337976)

[Table 8.12 Evaluation of Anchor Set (Common Items) Between 2018–19 and 2021–22 155](#_Toc138337977)

[Table 8.13 Linked Item Parameter Results 156](#_Toc138337978)

[Table 8.14 IRT Summary *b*-value Estimates for All CSA Operational Items 158](#_Toc138337979)

[Table 8.15 Item Difficulty Parameter Distribution by Grade Level or Grade Band 159](#_Toc138337980)

[Table 8.16 Conversion of Theta Score to Reporting Scores by Grade Level or Grade Band Based on the 2018–19 Test Administration (the Baseline) 161](#_Toc138337981)

[Table 8.17 Test Reliability of the Total Scores 164](#_Toc138337982)

[Table 8.18 Scale Score CSEMs at the Score Reporting Range Thresholds 167](#_Toc138337983)

[Table 8.19 Decision Accuracy for Reaching a Reporting Range 168](#_Toc138337984)

[Table 8.20 Decision Consistency for Reaching a Reporting Range 168](#_Toc138337985)

[Table 8.A.1 Item Difficulty Distributions by Domain 181](#_Toc138337986)

[Table 8.A.2 Item-Total Correlation Distributions by Domain 182](#_Toc138337987)

[Table 8.A.3 Item Difficulty, Item-Total Correlation, and Flag Status by Item for Grade   
Three 183](#_Toc138337988)

[Table 8.A.4 Item Difficulty and Item-Total Correlation by Item for Grade Four 185](#_Toc138337989)

[Table 8.A.5 Item Difficulty and Item-Total Correlation by Item for Grade Five 187](#_Toc138337990)

[Table 8.A.6 Item Difficulty and Item-Total Correlation by Item for Grade Six 189](#_Toc138337991)

[Table 8.A.7 Item Difficulty and Item-Total Correlation by Item for Grade Seven 191](#_Toc138337992)

[Table 8.A.8 Item Difficulty and Item-Total Correlation by Item for Grade Eight 193](#_Toc138337993)

[Table 8.A.9 Item Difficulty and Item-Total Correlation by Item for High School 195](#_Toc138337994)

[Table 8.B.1 IRT Item Difficulty for Grade Three 197](#_Toc138337995)

[Table 8.B.2 IRT Item Difficulty for Grade Four 199](#_Toc138337996)

[Table 8.B.3 IRT Item Difficulty for Grade Five 201](#_Toc138337997)

[Table 8.B.4 IRT Item Difficulty for Grade Six 203](#_Toc138337998)

[Table 8.B.5 IRT Item Difficulty for Grade Seven 205](#_Toc138337999)

[Table 8.B.6 IRT Item Difficulty for Grade Eight 207](#_Toc138338000)

[Table 8.B.7 IRT Item Difficulty for High School 209](#_Toc138338001)

[Table 8.C.1 Total Testing Time (in Minutes) at Each Raw Score Interval—Grade Three 211](#_Toc138338002)

[Table 8.C.2 Total Testing Time (in Minutes) at Each Raw Score Interval—Grade Four 212](#_Toc138338003)

[Table 8.C.3 Total Testing Time (in Minutes) at Each Raw Score Interval—Grade Five 212](#_Toc138338004)

[Table 8.C.4 Total Testing Time (in Minutes) at Each Raw Score Interval—Grade Six 213](#_Toc138338005)

[Table 8.C.5 Total Testing Time (in Minutes) at Each Raw Score Interval—Grade Seven 213](#_Toc138338006)

[Table 8.C.6 Total Testing Time (in Minutes) at Each Raw Score Interval—Grade Eight 214](#_Toc138338007)

[Table 8.C.7 Total Testing Time (in Minutes) at Each Raw Score Interval—High School, All Grade Levels 214](#_Toc138338008)

[Table 8.D.1 Reliabilities and SEMs by Gender 215](#_Toc138338009)

[Table 8.D.2 Reliabilities and SEMs by Economic Status 215](#_Toc138338010)

[Table 8.D.3 Reliabilities and SEMs by Special Education Services 216](#_Toc138338011)

[Table 8.D.4 Reliabilities and SEMs by Attendance in US Schools 216](#_Toc138338012)

[Table 8.D.5 Reliabilities and SEMs by Spanish Instruction Status 217](#_Toc138338013)

[Table 8.D.6 Reliabilities and SEMs by English Proficiency 217](#_Toc138338014)

[Table 8.D.7 Reliabilities and SEMs by English Proficiency (Continued) 218](#_Toc138338015)

[Table 8.E.1 Reliability of Classification for Grade Three: Decision Accuracy 219](#_Toc138338016)

[Table 8.E.2 Reliability of Classification for Grade Three: Decision Consistency 219](#_Toc138338017)

[Table 8.E.3 Reliability of Classification for Grade Four: Decision Accuracy 219](#_Toc138338018)

[Table 8.E.4 Reliability of Classification for Grade Four: Decision Consistency 219](#_Toc138338019)

[Table 8.E.5 Reliability of Classification for Grade Five: Decision Accuracy 220](#_Toc138338020)

[Table 8.E.6 Reliability of Classification for Grade Five: Decision Consistency 220](#_Toc138338021)

[Table 8.E.7 Reliability of Classification for Grade Six: Decision Accuracy 220](#_Toc138338022)

[Table 8.E.8 Reliability of Classification for Grade Six: Decision Consistency 220](#_Toc138338023)

[Table 8.E.9 Reliability of Classification for Grade Seven: Decision Accuracy 221](#_Toc138338024)

[Table 8.E.10 Reliability of Classification for Grade Seven: Decision Consistency 221](#_Toc138338025)

[Table 8.E.11 Reliability of Classification for Grade Eight: Decision Accuracy 221](#_Toc138338026)

[Table 8.E.12 Reliability of Classification for Grade Eight: Decision Consistency 221](#_Toc138338027)

[Table 8.E.13 Reliability of Classification for High School: Decision Accuracy 222](#_Toc138338028)

[Table 8.E.14 Reliability of Classification for High School: Decision Consistency 222](#_Toc138338029)

List of Figures

[Figure 3.1 Sample CSA practice test item 23](#_Toc138168081)

Acronyms and Initialisms Used in the *California Spanish Assessment Technical Report*

|  |  |
| --- | --- |
| Term | Definition |
| 1PL | one-parameter logistic |
| 1PL-IRT | one-parameter logistic item response theory |
| ADEL | adult English learner |
| AERA | American Educational Research Association |
| AIS | average item score |
| ALTRD | Assessment and Learning Technology Research & Development |
| APA | American Psychological Association |
| 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 |
| CCCSSeE | California Common Core State Standards en Español |
| *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 |
| CSA | California Spanish Assessment |
| CSEM | conditional standard error of measurement |
| *DFA* | *Directions for Administration* |
| DIF | differential item functioning |
| *EC* | *Education Code* |
| EL | English learner |
| ELA | English language arts/literacy |
| ELPAC | English Language Proficiency Assessments for California |
| eSKM | Enterprise Score Key Management |
| GPCM | generalized partial credit model |
| HOSS | highest obtainable scale score |
| IBIS | Item Banking Information System |
| IEP | individualized education program |
| IFEP | initial fluent English proficient |
| IRF | item response function |
| IRM | item review meeting |
| IRT | item response theory |
| ISAAP | Individual Student Assessment Accessibility Profile |
| K | kindergarten |

Table of Acronyms and Initialisms *(continuation)*

|  |  |
| --- | --- |
| Term | Definition |
| LEA | local educational agency |
| LOSS | lowest obtainable scale score |
| MC | multiple choice |
| MH | Mantel-Haenszel |
| NCME | National Council on Measurement in Education |
| OTI | Office of Testing Integrity |
| PAR | Psychometric Analysis & Research |
| PCM | partial credit model |
| QA | quality assurance |
| RFEP | reclassified fluent English proficient |
| RMSEA | root mean square error of approximation |
| RSD | ratio of standard deviations |
| SBE | State Board of Education |
| SCOE | Sacramento County Office of Education |
| SD | standard deviation |
| SEM | standard error of measurement |
| SFTP | secure file transfer protocol |
| SMD | standardized mean difference |
| SR | selected response |
| SSID | Statewide Student Identifier |
| SSR | Student Score Report |
| STAIRS | Security and Test Administration Incident Reporting System |
| TCC | test characteristic curve |
| TDS | test delivery system |
| TEI | technology-enhanced item |
| TIF | test information function |
| TOMS | Test Operations Management System |
| UAT | user acceptance testing |
| *USC* | *United States Code* |

## Introduction

This chapter provides an overview of the California Spanish Assessment (CSA), including background information, purpose of the assessment, intended population, testing window, organizations and systems involved in the development and administration of the CSA.

### 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.

During the 2021–22 test administration, the CAASPP System comprised the following assessments:

* Smarter Balanced assessments and tools:
* Summative Assessments—Computer-based assessments for English language arts/literacy (ELA) and mathematics in grades three through eight and grade eleven
* Interim Assessments—Optional resources developed for grades three through eight and grade eleven designed to inform and promote teaching and learning by providing information that can be used to monitor student progress toward mastery of the Common Core State Standards that may be administered to students at any grade level
* Tools for Teachers—Professional development materials and instructional resources designed to help teachers use formative assessment processes for improved teaching and learning in all grade levels
* 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, eight, and high school (grade ten, eleven, or twelve; these are the California Science Test and the CAA for Science)
* The CSA, optional for eligible students in grades three through eight and high school (grades nine through twelve) and designed to measure a student’s literacy in Spanish reading/language arts inclusive of reading, writing mechanics, and listening, with the intent to serve as a high school measure suitable to be used in part for the California State Seal of Biliteracy once all requirements are met

As part of the CAASPP System of assessments, the CSA was approved as an optional primary language and reading/language arts assessment in Spanish by the California State Board of Education (SBE).

Development of the CSA started in September 2016 with the SBE’s approval of the high-level test design. Following the 2018 fall CSA field test, the first CSA operational administration was administered optionally to students seeking a measure of their Spanish reading/language arts skills during spring 2019. The spring 2021–22 test administration marks the fourth operational year of the CSA, including the 2019–20 test administration, which was suspended in the middle of the testing window because of the novel coronavirus disease 2019 pandemic.

More background information about the CAASPP System can be found on the *CalEdFacts* web page on the California Department of Education (CDE) website.

### Test Purpose(s)

This computer-based assessment for students in grades three through eight and high school was designed to measure a student’s Spanish reading/language arts skills in reading, writing mechanics, and listening for the purposes of providing

* student-level data in Spanish literacy;
* aggregated data that may be used for evaluating the implementation of Spanish language arts programs at the local level; and
* a high school measure suitable to be used, in part, for the California State Seal of Biliteracy. Currently, the CSA does not meet the requirements identified in California *Education Code (EC)*, Section 51460(a) for the State Seal of Biliteracy because the speaking domain is not included in the assessment.

The CSA is aligned with the translated and linguistically augmented version of the Common Core State Standards for ELA, which are called the California Common Core State Standards en Español (CCCSSeE). The CSA measures a student’s Spanish reading/language arts skills in reading, writing mechanics, and listening based on the CCCSSeE.

### Test Design

The CSA is administered to students in grades three through twelve. The CSA was delivered online with fixed (nonadaptive) forms—one general form, one accommodated form, and one braille form. The general form was delivered in three versions, with each having the same operational items but different embedded field test items. The accommodated and braille forms had accommodated resources specific to the targeted population for these forms. All items were machine-scored.

### Intended Population

The population for the CSA comprises all California students in grades three through twelve who receive instruction in Spanish and who seek a measure that recognizes their Spanish-specific reading, writing mechanics, and listening skills. To be able to participate, students were assigned the CSA by an LEA CAASPP coordinator.

During the first operational CSA administration in 2018–19, the number of students taking the CSA varied significantly across different grade levels, from approximately 10,000 in grade three to fewer than 1,500 in grade twelve. This trend may have been affected by the emergence of Spanish-reading/language arts programs within local educational agencies (LEAs) and how LEAs exercised the option to assign students to take this optional assessment.

In the 2021–22 test administration, the number of students who completed the CSA increased in comparison with the 2019–20 and 2020–21 test administrations. In the 2021–‍22 test administration, the number of students who completed the test ranged from approximately 11,000 in grade three to 3,500 in high school, as indicated in table 8.2.

### Intended Use and Purpose of Test Scores

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

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

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

Therefore, the two primary purposes of a test within the CAASPP System are:

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 CSA for grades three through eight and high school are administered within a testing window pursuant to *California Code of Regulations*, Title 5 [5 *CCR*]*,* sections 855(a)(1), 855(a)(2), 855(b), and 855(c). The testing window starts in the middle of January and ends in the middle of July; however, LEAs can begin testing only once 66 percent of the LEA’s instructional year has been completed. For the 2021–22 CSA administration, the state testing window opened on January 11, 2022, and ended on July 15, 2022.

A student could take the CSA within the testing window over as many days as required to meet the student’s needs (5 *CCR,* Education, Division 1, Chapter 2, Subchapter 3.75, Article 2*,* Section 855[a]). Similar to other CAASPP assessments, the CSA was untimed for test takers. The estimated duration for the assessment was approximately two hours, on average. Refer to subsection [*2.2.1 Test Length*](#_Test_Length) for more information about test length.

### Impact of the Novel Coronavirus Disease 2019 Pandemic

#### Remote Testing Flexibility

In the 2021–22 administration, LEAs continued to use two options for testing students using the CAASPP:

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). Remote administration is intended as an option for an LEA only when its students are receiving remote instruction. Students receiving in-person instruction should be tested in person.

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

### Significant CSA Developments in 2021–22

#### Accessibility Resources

The following accessibility resource–related updates were made:

* The turn off any universal tools embedded designated support was indicated in the Test Administrator Interface; test administrators were notified during the test entry approval process that an individual student needed universal tools turned off.
* In-test widgets, such as help guides, glossaries, and calculators, were resizable and movable on-screen. This allowed students to move or resize widgets to increase focus on desired content.
* LEAs could upload supporting documents as part of the request for an unlisted resource and could use the Notes feature to communicate with the CDE.

#### Reused Forms for High School in 2021‒22

Because of the small testing sample for high school in the 2020‒21 test administration (262 students in total took the CSA for high school), not enough testing samples were collected to support the item analysis activities for items that were field-tested in 2020–21. Consequently, no new forms for high school could be developed for the 2021–22 test administration. All 2019–20 forms for high school, including the general forms, accommodated forms, and braille forms, as well as the embedded field test items, were reused during the 2020‒21 and 2021–22 test administrations.

The general forms for grades three through eight were refreshed forms, given the sufficient testing samples in the 2020‒21 test administration, while the accommodated forms and braille forms were reused from the 2019–20 test administration.

#### High School Postequating Procedures

Beginning in the 2018–19 test administration, the one-parameter logistic item response theory (IRT) measurement model was implemented in the CSA psychometric analyses for item calibration, equating, and scaling. For the 2020–21 test administration, IRT postequating procedures were used for grade three through grade eight but not for high school because of the insufficient testing sample (262 students). Instead, the classical linear equating method was used for the high school grade band to link the 2020–21 data to the 2018–19 base scale.

In the 2021‒22 test administration, the sufficient testing sample in high school made the one-parameter logistic IRT model feasible to be applied. As a result, the IRT postequating procedure was used for high school.

### Groups and Organizations Involved with the CAASPP System

#### 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 is also the state educational agency responsible for overseeing California’s compliance with programs that meet the requirements of the federal Every Student Succeeds Act 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, 2022c).

#### California Department of Education

The CDE oversees California’s public school system, which is responsible for the education of more than 5,800,000 children and young adults in more than 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, 2022b).

#### California Educators

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

#### Contractors

A number of organizations contribute to the success of the CSA.

##### Primary Testing Contractor—ETS

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

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

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

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

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

* providing the CAI proprietary TDS, including the Student Testing Interface, Test Administrator Interface, secure browser, and practice and training tests;
* hosting and providing support for its TDS, a component of the overall CAASPP Assessment Delivery System;
* 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 used by LEAs to manage all aspects of CAASPP testing. TOMS serves various functions, including, but not limited to, the following:

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

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

LEA staff involved in the administration of the CAASPP—such as LEA CAASPP coordinators, CAASPP test site coordinators, test administrators, and test examiners—are assigned varying levels of access to TOMS. For example, only an LEA CAASPP coordinator is given permission to set up the LEA’s test administration window; a test administrator or test examiner cannot download student reports. A description of user roles is explained more extensively in the *2021–22 CAASPP Online Test Administration Manual* (CDE, 2022a).

#### Test Delivery System

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

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 computer-based application through which the Student Testing Interface may be accessed. The secure browser prevents students from accessing other applications during testing.

#### Practice and Training Tests

All California testing programs have practice and training tests to inform educators, parents/‌guardians, and students about the individual tests. The practice and training tests were provided to LEAs to prepare students and LEA staff for administration of the CSA. These tests simulated the experience of the CSA computer-based assessments. Unlike the summative assessments, the practice and training tests did not gauge student success on the operational test, or 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. As remote testing was added for the 2021–22 administration, the practice and training tests permitted test administrators and students to practice using the remote monitoring and communication features.

The purposes of the training tests are to

* allow students and test administrators to quickly become familiar with the user interface and components of the TDS as well as with the process of starting and completing a testing session;
* allow students and test administrators to experience a grade-level assessment, grade-specific items and difficulty levels, and the format and structure of an operational assessment; 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.

#### California Educator Reporting System

CERS is the system used by LEAs to view preliminary student results from CAASPP testing. 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 or other characteristics of their choosing. The student results sent to CERS are appropriate for analysis of assessment results for use in informing instruction.

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

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

### Overview of the Technical Report

This technical report addresses the characteristics of the CSA administered in spring 2022 and contains nine additional chapters as follows:

* [Chapter 2](#_Overview_of_CSA) presents an overview of the processes involved in a testing cycle for the CSA. This chapter includes item development, test assembly, test administration, fairness and accessibility, generation of test scores, and psychometric analyses. The details on each stage in the testing process will be presented in the subsequent chapters.
* [Chapter 3](#_Item_Development_and_1) discusses the test design, blueprint, and item development for the 2021–‍22 test administration.
* [Chapter 4](#_Test_Assembly) discusses the content, psychometric criteria, and reviews that guide procedures of CSA test assembly.
* [Chapter 5](#_Test_Administration_1) details the processes involved in the administration of the CSA. It also describes the procedures followed by ETS to maintain test security throughout the test administration process.
* [Chapter 6](#_Standard_Setting) provides a brief summary of the standard setting process that established the base-year score reporting ranges.
* [Chapter 7](#_Scoring_and_Reporting) summarizes the types of scores and score reports that are produced at the end of each administration of the CSA.
* [Chapter 8](#_Psychometric_Analyses_1) summarizes the statistical procedures and results for 2021–22. These analyses include
* classical item analyses;
* test completion analyses;
* differential item functioning analyses;
* IRT calibration, postequating, and scaling;
* testing time analyses;
* reliability analyses; and
* validity analyses.
* [Chapter 9](#_Quality_Control_Procedures) highlights the quality control processes used at various stages of administration of the CSA.
* [Chapter 10](#_Continuous_and_Systematic_1) discusses the various procedures used to gather information to improve the CSA as well as strategies to implement possible improvements.

### Reference

*California* *Code of Regulations,* Title 5, Education, Division 1, Chapter 2, Subchapter 3.75, Article 2, Section 851.5. (n.d.).

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

California Department of Education. (2022a). *CAASPP online test administration manual.* Sacramento, CA: California Department of Education.

California Department of Education. (2022b, August). *Organization.* California Department of Education website.

California Department of Education. (2022c, October). *State Board of Education responsibilities.* California Department of Education website.

*California Education Code,* Title 2, Elementary and Secondary Education, Division 4, Part 28, Chapter 3, Article 6.

## Overview of CSA Processes

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

### Item Development

ETS used the 180 field test items developed in 2019–20 across the six grade levels (i.e., 30 items each for grades three through eight) for the 2021–22 test administration—the fourth operational administration—and delivered them to the California Department of Education (CDE) via the ETS Item Banking Information System (IBIS). The total number of machine-scorable items developed in 2019–20 (415) was greater than the number to be field-tested (180) in the 2021–22 test administration because end-of-contract overage was built in.

The developed items were designed to be engaging to the student population and represented a wide variety of item types. All items for the CSA field tests were developed in accordance with the *ETS Standards for Quality and Fairness* (ETS, 2014) across all phases of item and test development. While under initial development, the assessment materials, including items, passages, and listening stimuli, were stored on password-protected ETS computers and secure internal network drives. Audio recordings were produced as electronic audio files and delivered securely to the CDE for review.

All secure documents needed for CDE review that were not available in IBIS were delivered to the CDE via a secure file transfer protocol server.

#### Item Format

The CSA includes the following primary computer-based item formats:

* **Selected-response items—**Students are instructed to select one or more choices. Most CSA items have three or four options; a few items have more than four options.
* **Technology-enhanced items—**Technology beyond simple option selection is incorporated in some items.

Detailed information on item format is included in subsection [*3.1.6 Item Types*](#_Item_Types) in [*Chapter 3: Item Development and Review*](#_Item_Development_and_1). All items included in the 2021–22 CSA forms were machine-scorable.

#### Item Specifications

The CSA item specifications provide descriptions of item characteristics that are intended to measure each content standard consistently. They were developed based on the California Common Core State Standards en Español guidelines. During item development, item writers were provided CSA item specifications and a CSA style guide that contained detailed information about the consistency in item development and item review processes. Refer to subsection [*3.1.4 Item Specifications*](#_Item_Specifications_1) in [chapter 3](#_Item_Development_and_1) for detailed information about item specifications.

### Test Assembly

For the 2019–20 high school operational forms that were reused for 2021–22 CSA testing in high school grade band (i.e., grades nine through twelve), the ETS Assessment and Learning Technology Research & Development (ALTRD) team used 52 items that were previously administered during the 2018–19 operational administration and the fall 2018 field test. The operational forms for the 2021‒22 test administration in grades three through eight were refreshed forms. The CDE reviewed the 2021–22 forms in IBIS before they were configured in the test delivery system (TDS) by Cambium Assessment, Inc.

Note that test assembly activities were not conducted for the 2021–22 test administration in high school forms, because the sample size in the 2019–20 and 2020–21 test administrations were too small to conduct item analysis. As a result, the 2019–20 forms were reused in the 2021–‍22 test administration for high school only.

For grades three through eight, the 2021‒22 forms were assembled and not reused, and psychometric criteria were specified for the test form review before the test administration. The psychometric guidelines of item selection and form building were developed and shared with ALTRD in the previous spring, before the start of the test assembly process for the next test administration.

ETS content staff and Psychometric Analysis & Research staff reviewed the assembled forms thoroughly in regard to the following aspects of the operational forms:

* Coverage of blueprints
* Overall test design and statistical properties
* Statistical properties of individual items

Details of the psychometric criteria of form review are included in subsection [*4.3.3 Psychometric Criteria and Identification of Eligible Items*](#_Psychometric_Criteria_and_1).

#### Test Length

The CSA is a computer-based linear assessment. The general forms for grades three through eight contained 52 operational items and 10 field test items. The general form for high school contained 52 operational items and 16 field test items. The accommodated forms for all grade levels and the grade band contain 52 operational items and no field test items.

##### Operational Testing

While CSA operational items included a wide variety of item types, multiple-choice items made up approximately 80 percent of the overall pool of items on forms in 2021–22. Passage-based items were a central feature of the CSA; few items were discrete (stand-alone) items. Each grade level or the high school grade band had one general core of 52 operational items; unique field tests were embedded into each grade level or grade band’s core, yielding three distinct forms that were assigned to students randomly.

##### Field Testing

Field test items have been embedded into the CSA operational tests since the 2019–20 test administration. Ten field test items were embedded into each of the 2021–22 forms for grades three through eight, so each grade level’s test form was composed of 62 items. Similarly, 16 field test items were embedded into the high school 2021–22 operational form, so that test form was composed of 68 items. Additional information about CSA test assembly can be found in [*Chapter 4: Test Assembly*](#_Test_Assembly).

#### Test Blueprints

Blueprints represent a set of constraints and specifications to which each test form must conform. The CSA had three main subcontent areas or domains: reading, writing mechanics, and listening. The blueprints of the assessment are presented in table 4.A.1 through table 4.A.3 in [appendix 4.A](#_Appendix_4.A_Test).

#### Item Selection

Items are selected for placement on forms based on items’ exposure over time. A certain number of items, called anchor items, are repeated from a grade level or high school grade band’s previous form, and these items collectively share attributes of the whole form in regard to difficulty and coverage of the three domains. The other operational items may have been used before or may be new to the form.

The process of replacing a predetermined minimum number of items with each new form’s creation is known as a refresh. The ratio of replaced items to the overall number of operational items (52) is called the refresh rate.

Items selected for use will also need to meet psychometric criteria. For details, refer to subsection [*4.3.3 Psychometric Criteria and Identification of Eligible Items*](#_Psychometric_Criteria_and_1).

### Test Administration

The CSA was administered using the secure browser and TDS, ensuring a secure, confidential, standardized, consistent, and appropriate administration for students. Additional information about the administration of the CSA can be found in [*Chapter 5: Test Administration*](#_Test_Administration_1).

#### Test Security and Confidentiality

All operational tests within the California Assessment of Student Performance and Progress (CAASPP) System are secure. For the CSA administration, every person having 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 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’_Office_of_1)in [*Chapter 5: Test Administration*](#_Test_Administration_1).

In the pursuit of enforcing secure practices, ETS strives to safeguard the various processes involved in a test development and administration cycle. Those processes are listed next. The practices related to each of the following security processes are discussed in detail in section [*5.6 Test Security and Confidentiality*](#_Test_Security_and_1):

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

#### Procedures to Maintain Standardization

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

The CSA is administered in conjunction with the other assessments that compose the CAASPP System. ETS employs processes to ensure the standardization of an administration cycle; these processes are discussed in more detail in section [*5.2 User Roles and Standardization*](#_User_Roles_and_1).

Staff at local educational agencies (LEAs) involved in the CAASPP 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* (CDE, 2022b).

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

* ***CAASPP Online Test Administration Manual*—**This web-based manual provides test administration procedures and guidelines for LEA CAASPP coordinators and 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, 2022b). (Refer to [*5.2.4.2 CAASPP Online Test Administration Manual*](#_CAASPP_Online_Test)in [chapter 5](#_Test_Administration_1) for more information.)
* ***CAASPP and English Language Proficiency Assessments for California (ELPAC) Test Operations Management System (TOMS) User Guide*—**This web-based manual provides instructions for TOMS, allowing LEA staff, including LEA CAASPP coordinators and CAASPP test site coordinators, to perform several tasks, including setting up test administrations, adding and managing users, assigning tests, and configuring computer-based student test settings (CDE, 2022a). (Refer to [*5.2.4.3 CAASPP and ELPAC Test Operations Management System User Guide*](#_CAASPP_and_English_1)in [chapter 5](#_Test_Administration_1) for more information.)

### Fairness and Accessibility

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

#### Overview

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

#### 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, 2021).

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

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

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

**Unlisted resources** are non-embedded and made available if specified in the eligible student’s IEP or Section 504 plan and only on approval by the CDE.

[Appendix 5.A](#_Appendix_5.A:_Accessibility) presents counts and percentages of students assigned designated supports, accommodations, and unlisted resources for the 2021–22 CSA administration. The tables in [appendix 5.A](#_Appendix_5.A:_Accessibility) were created using student demographic data in version 2 of the production data file (“P2”) updated on August 21, 2022.

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

#### Description of Differential Item Functioning

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

### Scores

Individual student scores were reported for the 2021–22 CSA administration. Student performance on the reporting scale was designated into one of the three reporting ranges presented in subsection [*7.1.3 Score Reporting Ranges*](#_Score_Reporting_Range). For information regarding score specifications and score reports, refer to [*Chapter 7: Scoring and Reporting*](#_Scoring_and_Reporting).

#### Score Reporting

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

CSA scores can also be viewed through the California Educator Reporting System (CERS), a secure website that provided authorized users with interactive and cumulative online reports for content area at the student, school, and LEA levels. CERS also provided individual score reports. Refer to subsection [*7.3.1 Online Reporting*](#_Online_Reporting_1) for details about TOMS and CERS and subsection[*7.3.3 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, CSA 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 at the state, LEA or direct funded charter school, or school level; combining student scores for all students; or by combining student scores for students who represent selected demographic student groups.

The aggregation procedures used to present CSA results are described in section [*7.2 Overview of Score Aggregation Procedures*](#_Overview_of_Score). Aggregated *r*esults by demographic variables are presented in [appendix 7.C](#_Appendix_7.C:_Demographic_1). In table 7.C.1 through table 7.C.7, 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 reporting range. To protect student privacy, statistics are presented in the tables as “N/A” when the number of students in the sample is 10 or fewer. Definitions for the demographic student groups included in these tables are provided in table 7.5.

### Psychometric Analyses

Psychometric analyses were conducted on the data from the CSA, including classical item analyses, DIF analyses, item response theory (IRT) calibration and linking, testing time analyses, and reliability analyses. The results of these analyses support understanding of item performances and internal structure of the assessment and provide validity evidence for both response processes and scoring. Detailed descriptions of these analyses are presented in [*Chapter 8: Psychometric Analyses*](#_Psychometric_Analyses_1).

#### Description of Classical Item Analyses

The psychometric analyses for the CSA 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 [*8.2 Classical Item Analyses*](#_Classical_Item_Analyses)*.* A description of the DIF analyses procedure is provided in section [*8.3 Differential Item Functioning Analyses*](#_Differential_Item_Functioning)*.*

#### Description of Item Response Theory Analyses

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

### References

California Department of Education. (2021). *California assessment accessibility resources matrix*. California Department of Education website.

California Department of Education. (2022a). *CAASPP and ELPAC Test Operations Management System user guide*. Sacramento, CA: California Department of Education.

California Department of Education. (2022b). *CAASPP online test administration manual*. Sacramento, CA: California Department of Education.

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

## Item Development and Review

This chapter describes the detailed procedures of item development and test form assembly for the California Spanish Assessment (CSA). Those processes include the ones that are entirely internal to ETS and those that are conducted in coordination with the California Department of Education (CDE) and Cambium Assessment, Inc. (CAI). In particular, development of new item types and features that differ from traditional item types are described.

CSA item development incorporated innovations and best practices from national language assessments. California Spanish-language teachers assisted in creating these items, and item review meetings (IRMs) with California educators were instrumental in determining the complexity of content and grade-level appropriateness. This section describes how California educators were selected and the process used to develop new items.

### Guidelines

Each item for the CSA was developed through a comprehensive development cycle and designed to conform to principles of item writing defined by ETS. Each item in the CSA operational item bank was developed to measure a specific California Common Core State Standards en Español (CCCSSeE) (San Diego County Office of Education, n.d.). In addition, guidelines for style, fairness, and bias and sensitivity helped item developers and reviewers to ensure consistency across the item development process.

#### Overview

While the 2021–22 high school forms were reused from the 2019–20 test administration, ETS chose 312 items that were from the 2018–19 forms or previously field-tested items for use on operational assessments across the other six grade levels—52 items on each general form for grades three through eight—and submitted them to the CDE by way of the ETS Item Banking Information System (IBIS). For the second time, approximately 20 percent of the items were refreshed for grades three through eight. This allowed items that were field-tested in fall 2018 or later to be used operationally for the first time.

All CSA items were designed to be engaging to the student population and represented a wide variety of item types. All items for the operational CSA were developed in accordance with the *ETS Standards for Quality and Fairness* (ETS, 2014) across all phases of item and test development. While under initial development, the assessment materials, including items, passages, and listening stimuli, were stored on password-protected ETS computers and secure internal network drives. Audio recordings were produced as electronic audio files and delivered securely to the CDE for review.

All secure documents needed for CDE review that were not available in IBIS were delivered to the CDE via the secure file transfer protocol server. Table 3.1 presents the number of items per form.

#### Plan

The items developed for use as operational and field test items in the 2021–22 CSA forms closely reflected the distribution of domains in the blueprint (CDE, 2017). The total number of machine-scorable items developed (415) was greater than the number used as field test items (228) in the CSA 2021–22 forms because end-of-contract overage was built in to meet contractual obligations. ETS develops overage (i.e., more items than the number needed as determined by the refresh rates) every year to account for the potential rejection of items during item review and data review meetings. If item reviewers at the IRM determined that certain items were not appropriate for operational testing, the overage ensured that the minimum item counts for the assessment forms would be satisfied.

For the general assessment forms, there was substantial overage built in. However, for the accommodated forms, there was little overage when accommodations, such as text-to-speech, closed-captioning, audio transcript, and braille were applied to the items. In contrast, the current policy is for every item testing a standard that is eligible to be tested on accommodated forms to be created (“born”) accessible, removing the need to later apply accommodations to such items.

Table 3.1 shows the number of items originally developed for the 2020–21 test administration and field-tested in each of the domains of reading, writing mechanics, and listening for the CSA in the 2021–22 test administration, one year later than anticipated because of the aftermath of the novel coronavirus disease 2019 (COVID-19) pandemic. The total number of items developed across the six grade levels and the high school grade band (i.e., grades nine through twelve) was 415.

Table 3.1 **Number of Items Developed per Grade Level and Grade Band for the CSA**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Domain** | **Grade 3** | **Grade 4** | **Grade 5** | **Grade 6** | **Grade 7** | **Grade 8** | **High School** |
| Listening | 8 | 8 | 8 | 8 | 8 | 8 | 16 |
| Reading | 27 | 27 | 27 | 27 | 27 | 27 | 39 |
| Writing Mechanics | 18 | 18 | 18 | 18 | 18 | 18 | 42 |
| **Total:** | **53** | **53** | **53** | **53** | **53** | **53** | **97** |

All items created for the CSA adhere to the *ETS* *Standards for Quality and Fairness* (2014)across all phases of item and test development. Each CSA item was developed through a comprehensive development cycle and designed to conform to the principles of quality item writing as defined by ETS.

#### Process

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 or set of correct answers (for selected-response items only).
5. Distractors are plausible but are incorrect or do not correctly answer the stem (for selected response only).
6. The answer key is correct.
7. Item format and content adhere to the principles of universal design.

#### Item Specifications

ETS maintained item specifications for the CSA that describe the characteristics of items written to measure the CCCSSeE that, in turn, provide evidence for the CSA’s reading, writing mechanics, and listening domains. Using the item specifications helped ensure that all items developed for the CSA measured standards consistently. Item writing assignments were guided by the CSA blueprints, developed in consultation with the CDE (CDE, 2020).

The specifications include

* a description of best practices for item writing:
* universal design,
* bias and sensitivity avoidance,
* cognitive level,
* anatomy of an item,
* item types and characteristics,
* a general list of elements to avoid, and
* stand-alone items;
* information about passages used to assess CSA domains;
* a description of standards used for items associated with reading passages, writing mechanics passages, and listening passages;
* a full statement of each standard featured on the CSA blueprint; and
* sample item stems at each grade level or the high school grade band for some standards.

#### Item Format

CSA items were developed with the understanding that students who are able may select responses using a mouse, touchscreen, or other supported input device. The majority of items were presented in a split-screen format, with a “stimulus” on the left side of the screen and the item to be answered on the right. The stimulus was usually a passage. This is shown in figure 3.1.



Figure 3.1 Sample CSA practice test item

A selected number of items had a multimedia stimulus, either a short audio file, or, for students with visual impairment, alternative text.

Items developed for the CSA were scored as being worth one point or two points.

#### Item Types

ETS developed a variety of technology-enhanced item (TEI) types that required the student respond to a question in different ways from typical selected-response items. Items could contain a stimulus (e.g., a passage, audio, or image).

Students responded to TEIs by typing an answer, completing a grid, dragging a response to a designated area, using a drop-down list selection, or selecting multiple areas in a graphic (also known as “hot spots”). As currently configured, all CSA item types were machine-scored in the test delivery system; there was no human scoring of CSA test questions.

Table 3.2 lists item types used in the CSA. Response types indicated with an asterisk (\*) are TEIs.

Table 3.2 **Item Types for the CSA**

|  |  |  |
| --- | --- | --- |
| **Response Type** | **Item Type** | **Description** |
| Multiple choice (MC) single select | MC | The item generally consists of a stem and list of choices; the test taker can select only one choice to respond. This may also include a stimulus. |
| MC multiple select | MC | The item generally consists of a stem and list of choices; the test taker can select two or more choices to respond. This may also include a stimulus. |
| Zones single select\* | Hot Spot | This is an item where the answer choices are predefined “hotspots” on an image. When the test taker selects (clicks) on the spot, the selection is highlighted, shaded, or outlined in red. The test taker selects one zone to respond. |
| Zone multiple select\* | Hot Spot | This is an item where the answer choices are predefined “hotspots” on an image. When the test taker selects (clicks) on the spot, the selection is highlighted, shaded, or outlined in red. The test taker selects two or more zones to respond. |
| Inline choice list single select\* | MC | The stem contains a single blank, and the test taker must fill the blank by selecting a choice from its corresponding choice list. |
| Inline choice list multiple select\* | MC | The stem contains two or more blanks, and the test taker must fill each blank by selecting a choice from the corresponding choice lists. |
| Text choices single select\* | MC | The test taker responds by selecting only one of several underlined words or phrases embedded in a larger section of text. |
| Text choices multiple select\* | MC | The test taker responds by selecting two or more underlined words or phrases embedded in a larger section of text. |
| Grid multiple select\* | MC | The test taker responds by marking two or more cells in a table grid. |
| Match single select\* | Drag & Drop | The test taker responds by dragging and dropping a single choice (“source”) into the appropriate location (“target”). |
| Match multiple select\* | Drag & Drop | The test taker responds by dragging and dropping one or more choices (“sources”) into the appropriate locations (“targets”). |
| Composite\* | All item types except CR | The test taker completes multiple tasks based on a combination of machine-scored items. |

#### Item Banking

The CSA forms administered in 2019–20 were reused during the 2021–22 test administration for high school; other grade levels were refreshed. Following a typical operational administration of the CSA, the operational forms across all grade levels and the high school grade band are refreshed for future test administrations. To support the proposed refresh rates of 20 percent for grades three through eight and 35–50 percent for high school, it is necessary to build an item bank where the content and statistical attributes of each item are included. All the items in the item bank need to be calibrated and linked onto the base scales that were established after the CSA’s first operational administration, which occurred during the 2018–19 California Assessment of Student Performance and Progress (CAASPP) administration.

The test forms used for the 2021–22 CSA administration included operational and field test items. As a standardized procedure, after each CSA administration, initial item analyses are implemented, and the results are reviewed by ETS Psychometric Analysis & Research and Assessment and Learning Technology Research & Development staff, who provided recommendations to the CDE on whether the items should be included or excluded from the calibrations. Decisions were made in consultation with the CDE; details of this process are in section [*8.2 Classical Item Analyses*](#_Classical_Item_Analyses).

Next, the operational items are calibrated and postequated to the base scales established in the 2018–‍19 test administration. Refer to section [*8.4 Item Response Theory Analyses*](#_Item_Response_Theory) for calibration and linking. Final item analyses are usually conducted following the calibration and linking step after the testing window is closed.

Content experts from ETS and the CDE, as well as selected California educators, review the associated item statistics and evaluate the performance of items during the annual data review meeting. Particularly, they pay attention to reviewing the flagged items—those whose statistics fell beyond expected ranges—and work to provide plausible explanations for these particular items based on their knowledge of the student population.

With the CDE’s approval, the operational items and field test items are linked to the base scale through the process of item calibration and equating, as described in subsection [*8.4.3 Equating*](#_Equating_1) and subsection [*8.4.4 Calibration and Linking for the Field Test Items*](#_Calibration_and_Linking). Those equated operational items and embedded field test items, with their statistical information, are entered into the item bank for form assembly for future test administrations. It is expected that more new items will be developed, field-tested, and entered into the item bank for future test administrations. Over time, the item bank will expand gradually to support the rate of refresh.

#### Recruitment and Selection of Item Writers

Senior ETS content staff screened applications for item writers for the CSA, 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 used to assess Spanish reading/language arts in California. All item writers met the following minimum qualifications:

* Possession of a bachelor’s degree in a relevant field of education (An advanced degree in the relevant content was desirable.)
* Biliterate and bilingual (Spanish and English)

#### 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 CSA, ETS recruited and trained educators in the CCCSSeE.

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

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

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

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

### ETS Item Review Process

After items were drafted, ETS placed items developed for the CSA 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 CSA items.

#### Overview

Once an item was accepted for authoring, ETS employed a series of internal reviews. These reviews used established criteria to judge the quality of item content and to ensure that each item measured what it was intended to measure. These internal reviews also examined the overall quality of the items ahead of their being reviewed by the CDE and by educators at IRMs, which are described in more detail in section [*3.4 California Educator Review*](#_California_Educator_Review).

All items were entered into the IBIS with corresponding artwork and metadata. Within IBIS, items received ETS internal content, fairness, and editorial reviews.

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

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

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

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

#### ETS Content Review

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

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

Assessment specialists verified the classification of each item, both to evaluate the correctness of the classification and to confirm that the task posed by the item was relevant to the outcome it was intended to measure. The reviewers could accept the item and classification as written, suggest revisions, or recommend that the item be discarded. These steps occurred prior to 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 guided the review of all items, with a focus on accessibility for all student populations, and provided potential refinement solutions to improve the accessibility.

#### ETS Editorial Review

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

#### ETS Sensitivity and Fairness Review

ETS assessment specialists who were specially trained to identify and edit or eliminate questions that contained 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 and 2016). These trained staff members reviewed every item before the CDE and IRMs. Newly developed items were then submitted to the CDE for review prior to educator reviews.

The review process promoted a general 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 language learners of diverse backgrounds

### California Department of Education Review

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

### California Educator Review

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 CSA. The California educators fill an advisory role to the CDE and ETS and provide guidance on matters related to item development for the CSA.

However, because of the COVID-19 pandemic, the suspension of testing in 2019–20, and the decision prior to the conclusion of the item development process to reuse test forms from 2019–20 for the 2020–21 test administration, two educator review meetings were postponed and held virtually: May 2020 to October 2020 and May 2021 to October 2021.

ETS convened both meetings with California educators to

* review Spanish passages and items for grade-level appropriateness, content, bias and sensitivity, readability, and overall interest for the test taker; and
* obtain feedback from California educators about the passages and items to inform ETS on the appropriateness of their use on future test forms for the CSA.

Passages and items reviewed in October 2020 were moved to field test slots for the 2021–‍22 test administration, and passages and items reviewed in October 2021 were moved to field test slots for the 2022–23 test administration.

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

#### Composition of Item Review Panels

The panelists for the CSA IRM included current and former teachers, resource specialists, administrators, curriculum experts, and other education professionals. To qualify to participate in the IRM, educators had to self-assess their written and spoken Spanish as fluent. Preferred qualifications included

* currently being assigned to teach Spanish reading/language arts;
* currently working in Advanced Placement Spanish Language and Culture, on an International Baccalaureate in Spanish, or both;
* currently working in dual immersion or bilingual programs;
* currently serving heritage speakers;
* having a Spanish focus in the postsecondary studies background; and
* having studied or taught in a Spanish-speaking country.

Every effort was made to ensure that groups of item reviewers included a wide representation of genders, geographic regions, and ethnic groups in California. Efforts also were made to ensure representation by members with experience serving California’s diverse Spanish-learning population.

Table 3.3 shows the educational qualifications; present, self-reported occupation; and credentials of the individuals who participated in CSA IRMs.

Table 3.3 CSA Item Reviewer Qualifications

|  |  |  |
| --- | --- | --- |
| **Qualification Type** | **Qualification** | **Total** |
| **Occupation** | Spanish Teacher | 13 |
| **Occupation** | General Education Teacher | 11 |
| **Highest Degree Earned** | Bachelor’s Degree | 3 |
| **Highest Degree Earned** | Master’s Degree | 11 |

Table 3.3 *(continuation)*

|  |  |  |
| --- | --- | --- |
| **Qualification Type** | **Qualification** | **Total** |
| **Kindergarten (K)–12 Teaching Credential** | Elementary Teaching (multiple subjects) | 9 |
| **K–12 Teaching Credential** | Secondary Teaching (single subject) | 7 |
| **K–12 Teaching Credential** | Spanish | 7 |
| **K–12 Teaching Credential** | English Learner (CLAD, BCLAD) | 2 |
| **K–12 Teaching Credential** | Administrative | 0 |
| **K–12 Teaching Credential** | Other | 3 |

During a typical item review process, item reviewers are recruited through an online application process. Recommendations are solicited from local educational agencies and county offices of education as well as from CDE and California State Board of Education 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.

Table 3.4 provides the status of the items after the October 2020 IRM. None were rejected.

Table 3.4 Status of Items After the 2020 IRM

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Approved As Is** | **Approved with Revisions** | **Total Number of Items (Including Passages)** |
| Grade 3 | 39 | 14 | 57 |
| Grade 4 | 38 | 15 | 57 |
| Grade 5 | 38 | 15 | 57 |
| Grade 6 | 27 | 26 | 57 |
| Grade 7 | 39 | 14 | 57 |
| Grade 8 | 42 | 11 | 57 |
| High school | 87 | 10 | 104 |
| **Total:** | 310 | 105 | 446 |

Table 3.5 provides the status of the items after the October 2021 IRM. None were rejected.

Table 3.5 Status of Items After the 2021 IRM

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Approved As Is** | **Approved with Revisions** | **Total Number of Items (Including Passages)** |
| Grade 3 | 48 | 6 | 58 |
| Grade 4 | 40 | 13 | 57 |
| Grade 5 | 45 | 8 | 57 |
| Grade 6 | 44 | 9 | 57 |
| Grade 7 | 42 | 11 | 57 |
| Grade 8 | 45 | 8 | 57 |
| High school | 81 | 16 | 104 |
| **Total:** | 345 | 71 | 447 |

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

Repercussions from the COVID-19 pandemic precluded ETS from facilitating a traditional IRM to support the 2021–22 test administration.

Typically, ETS content-area assessment specialists facilitate CSA IRMs. 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 CSA
* Overview of the CSA 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 construct being assessed by the standard
* Difficulty range
* Clarity
* Correctness of the answer
* Plausibility of the distractors
* Bias and sensitivity factors

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

* Does the item
* have one and only one clearly correct answer or set of correct answers?
* measure the achievement standard?
* align with the construct being measured?
* test worthwhile concepts or information?
* Is the stimulus, if any, for the item
* required to answer the item?
* likely to be interesting to students?
* clearly and correctly labeled?
* providing all the information needed to answer the item?

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

### Data Review Meeting

After items were included in a field test slot administered to students, ETS prepared items that did not meet all statistical thresholds for review by the CDE and California educators. Review materials included field test items with their statistical data and statistical flags based on the respective test administration’s item analyses and provided the means to capture comments in the IBIS Content Review Tool.

Educators who were part of the data review panel were assigned a training video in Upskill—a centralized, online location for training materials—to give them an overview of what is involved in a DRM as well as an understanding of the statistical measures used to review the field test items. ETS conducted an introductory training to highlight any new issues and serve as a statistical refresher at the beginning of the data review meeting. 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 3.6 for the results of the item data review conducted on June 14 to June 15, 2022, for the 2021–22 test administration. This table shows the number of flagged items accepted without edits and the number of items rejected outright. The rejection rates were below 5 percent for each grade level or the high school grade band.

Table 3.6 Item Data Review Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Accept As Is** | **Reject** | **Total Items** | **Rejection Rate** |
| Grade 3 | 33 | 1 | 34 | 3% |
| Grade 4 | 33 | 1 | 34 | 3% |
| Grade 5 | 32 | 1 | 33 | 3% |
| Grade 6 | 34 | 0 | 34 | 0% |
| Grade 7 | 34 | 0 | 34 | 0% |
| Grade 8 | 34 | 0 | 34 | 0% |
| High school | 51 | 0 | 51 | 0% |

### References

California Department of Education. (2017). *Draft California Spanish Assessment blueprint.* California Department of Education website.

California Department of Education. (2020). *Item content specifications for the California Spanish Assessment.* [Unpublished manuscript]. Sacramento, CA: California Department of Education.

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

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

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

San Diego County Office of Education. (n.d.) *CCSS en Español: Common Core State Standards*. San Diego County Office of Education website.

## Test Assembly

This chapter discusses the detailed procedures of test assembly for the 2021–22 California Spanish Assessment (CSA) administration.

### Test Design

The CSA design is based on the California State Board of Education–approved high-level test design for an operational assessment, which requires that students in the tested grade levels or the high school grade band (grades nine through twelve) provide responses to a fixed form of 52 operational items (California Department of Education [CDE], 2016). The CSA is administered to students in grades three through twelve. The CSA was delivered online with fixed forms—one general form, one accommodated form, and one braille form. The general form was delivered in three versions, with each having the same operational items but different embedded field test items. The accommodated and braille forms had accommodated resources specific to the targeted population for these forms. All items were machine-scored.

The braille forms and accommodated forms did not include field test items because of the increase in testing time needed by the students using those resources.

All general forms for grades three through eight were newly developed forms. The high school general forms were reused from the 2019–20 test administration.

### Test Blueprints and Other Content Specifications

#### Test Blueprints

Each operational assessment form contained items that met the proportions in the test blueprint (CDE, 2017). The test blueprint for the CSA provides the proposed number of items to be included in an operational assessment for each Spanish reading/language arts domain assessed in grades three through eight and high school (CDE, 2017).

Table 4.1 shows the distribution of the operational assessment items by domain and grade level or the high school grade band. [Appendix 4.A](#_Appendix_4.A_Test) presents the overview of the CSA blueprint by grade span.

Table 4.1 Number of CSA Items to Administer per Form

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Domain** | **Grade 3** | **Grade 4** | **Grade 5** | **Grade 6** | **Grade 7** | **Grade 8** | **High School** |
| Listening | 12 | 12 | 12 | 12 | 12 | 12 | 12 |
| Reading | 24 | 24 | 24 | 24 | 24 | 24 | 24 |
| Writing Mechanics | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| **Total Items:** | 52 | 52 | 52 | 52 | 52 | 52 | 52 |

### Test Production Process

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

#### Content Criteria

Operational items in 2021–22 for high school were identical to those assembled for the suspended 2019–20 CSA administration, but the forms for grades three through eight were refreshed for administration in 2021–22. Field test items were added to the forms for grades three through eight. For more information on how the additional field test items were selected to be added, refer to subsection [*3.4.2 Meetings for Review of CSA Field Test Item*s](#_Meetings_for_Review).

For grades three through eight, test developers selected items from the eligible item pool, ensuring that, as a whole, test forms

* met the coverage specifications of the test blueprints (subsection [*4.2.1 Test Blueprints*](#_Test_Blueprints)),
* met item selection criteria developed by the ETS psychometrics team (subsection [*4.3.3 Psychometric Criteria and Identification of Eligible Items*](#_Psychometric_Criteria_and_1)),
* represented a variety of item types, and
* provided a wide variety of item context.

#### Test Forms

In addition to the forms for the general administration, there was also one computer-based braille form and one accommodated form available in the 2021–22 test administration.

All CSA forms were evaluated using the psychometric criteria as described in the following subsection.

#### Psychometric Criteria and Identification of Eligible Items

In addition to the CSA blueprint, statistical guidelines were developed by the ETS Psychometric Analysis & Research team to assist in test assembly. The guidelines included the following:

* All items were operationally ready, with item statistics.
* All items conformed to the specifications in the test blueprint.
* Items with *p*-values between 0.2 and 0.95 were used.
* Items that were too difficult or too easy—indicated by low or high *p*-values—were not used, as they served little purpose in evaluating test takers’ abilities.
* For polytomous items with a maximum of more than one point, the *p*-values could be obtained by dividing the average item score by the maximum score points.
* Items with polyserial correlations greater than 0.2 were to be used. However, given the limited number of CSA items in the item bank, for the 2021–22 test administration, a few items with slightly lower than 0.2 polyserial correlations were included to ensure complete test content coverage. Such items passed the data review and were confirmed with no content issues. The lower polyserial correlations might be due to the small sample size, with limited variance from the field test.
* Category C (large) differential item functioning (DIF) items were not to be included in the operational form.
* If, for content coverage reasons, it was necessary to include C-DIF items in the form, those items were reviewed by a DIF panel that included members of the focal groups that were affected. The members of the panel confirmed that the items were not biased. The panelists were not to have a vested interest in the outcome of the decision.
* If C-DIF items were to be selected, then a balance with regard to the direction of the C-DIF items was considered; that is, not all C-DIF items should be C- or C+ items.
* The CDE needed to sign off on any C-DIF items before they appeared on a test. Refer to section [*8.3 Differential Item Functioning Analyses*](#_Differential_Item_Functioning) for additional information about this criterion.
* Items with *b*-parameters were in the range of ±4.0.
* Items with standard error of *b*-parameters were 0.4 or less.

Because the one-parameter item response theory model was applied to the CSA, the *a-‍*parameterwas fixed without estimating.

Table 4.2 presents the number of forms and items that underwent psychometric review by grade level or the high school grade band.

Table 4.2 Number of Forms and Items Reviewed Psychometrically

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **N General Form** | **N Accommodated Form** | **N Operational Items in General Form** | **N Operational Items in Accommodated Form** | **N Field Test Items** | **Total Items (Operational + Field Test)** |
| Grade 3 | 3 | 2 | 52 | 52 | 30 | 111 |
| Grade 4 | 3 | 2 | 52 | 52 | 30 | 123 |
| Grade 5 | 3 | 2 | 52 | 52 | 30 | 122 |
| Grade 6 | 3 | 2 | 52 | 52 | 30 | 125 |
| Grade 7 | 3 | 2 | 52 | 52 | 30 | 123 |
| Grade 8 | 3 | 2 | 52 | 52 | 30 | 127 |
| High school | 3 | 2 | 52 | 52 | 46 | 139 |
| **Overall:** | **21** | **14** | **364** | **364** | **226** | **870** |

#### 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. 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 California Common Core State Standards en Español
* Verification of coverage of the blueprint
* Identification of any possible grammatical or production errors

#### California Department of Education Forms Review

Following the ETS content review, all proposed assessments were sent to the CDE for review to ensure the proposed assessments met CSA blueprint requirements and to verify there was no clueing between items or statistical issues. The CDE was provided with the following materials:

* Access to items in the item banking system
* A cover sheet indicating key details for the CDE review of the proposed forms in the Item Banking Information System
* Modified form planners
* Comment sheets

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

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

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

Comments from the CDE were resolved during a virtual meeting with 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 TDS. Unlike other stages of the test production process, this stage must occur prior to every administration of the CSA, even in the case of a form reuse.

Each item underwent an extensive platform review on different operating systems, such as Windows, Linux, and iOS, to ensure that the item 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 CSA UAT before the test could be released for administration to students.

#### Delivery of General Forms

CSA forms are fixed forms. Students receiving a general form were given the same operational items by grade level or grade band and version. Upon logging on and starting a testing opportunity for the first time, the student was randomly assigned a test form with unique embedded field test items available for the student’s grade level. Refer to subsection [*4.1 Test Design*](#_Test_Design) for discussion of general forms.

### Special Version Forms

#### Braille Form

ETS designed a braille form for students with visual impairment for the CSA. Items that were used for embedded designated supports and accommodations were also used for the braille form.

The items appeared in the same or similar positions on the braille form as they did in the general form for the general population.

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 or screen readers. Adaptation did not change the item type. Twinning an item meant the item used 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.

#### Accommodated Forms Other Than Braille

ETS designed a separate accommodated form to provide accessible content for those students who had one or more designated support or accommodation assigned, as determined by an educator, individualized education program team, or Section 504 plan. Items were embedded with content for text-to-speech and glossing of certain terms. Refer to [*5.4.1 Accessibility Resource Categories*](#_Accessibility_Resource_Categories_1) for a list of designated supports and accommodations available during the 2021–22 CSA administration. Items in accommodated forms were similar to, but not exactly the same as, those in general forms.

### References

California Department of Education. (2016). *Proposed high-level test design for the California Spanish Assessment.* California Department of Education website.

California Department of Education. (2017). *Draft California Spanish Assessment blueprint.* California Department of Education website.

California Department of Education. (2021). *Form assembly specifications for the 2021–‍2022 California Assessment of Student Performance and Progress California Spanish Assessment.* [Unpublished manuscript]. Sacramento, CA: California Department of Education.

### Appendix 4.A: Test Blueprints

The test blueprint for the CSA provides the proposed numbers of items and points to be included in an operational assessment for each Spanish reading/language-arts domain assessed in grades three through eight and high school.

All items are aligned with the translated and linguistically augmented version of the CCCSSeE.

Each grade level has more than 50 testable standards at its disposal. There are three overview tables provided—grades three through five, grades six through eight, and high school (grades nine through twelve)—to clarify the overall proportions of the blueprint. After the overview tables, specifics for each tested grade level are given, enumerating further subdivisions of the content and specific groups of testing standards. Note that high school grade levels are tested together in one level using the CCCSSeE designated as “9–10” and “11–‍12” and uses the designation “high school” (HS).

The blueprint is represented in tables. Each overview table is organized by the three domains assessed: reading, writing (mechanics), and listening—referred to as claim/score reporting category—and are provided in the first column. Other columns in the blueprint are as follows:

* *Second column:* Content category
* *Third column:* Number of items representing the content category on an operational assessment
* *Fourth column:* Number of points for the given content category
* *Remaining columns:* Aggregated item counts, points, and percentages by claim

Item counts and point values may be adjusted further during future stages of the CSA design and development effort to take into consideration the evaluation of pilot test results as well as the analyses of statistics of both the first field test and the first operational administration of the CSA.

Finally, for all tables, some items are anticipated to be polytomously scored (maximum of two points), so the number of items is smaller than the number of score points.

Table 4.A.1 CSA Test Blueprint for Grade Span Three Through Five

| **Claim/Score Reporting Category** | **Content Category** | **Total Items by Content Category** | **Total Score Points by Content Category** | **Total Items by Claim** | **Percentage of Items by Claim** | **Total Score Points by Claim** | **Percentage of Score Points by Claim** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Reading Claim:** Students can read, analyze, and interpret a variety of texts and genres through Spanish. | Literary | 6–9 | 7–11 | 24 | 46% | 27–35 | 40–58% |
| **Reading Claim:** Students can read, analyze, and interpret a variety of texts and genres through Spanish. | Informational | 6–9 | 7–11 | 24 | 46% | 27–35 | 40–58% |
| **Reading Claim:** Students can read, analyze, and interpret a variety of texts and genres through Spanish. | Vocabulary and Meaning | 8–10 | 10–13 | 24 | 46% | 27–35 | 40–58% |
| **Writing Mechanics Claim:** Students can revise writing products that accurately and convincingly present, describe, and explain ideas for a range of purposes and audiences through Spanish. | Foundational Mechanics and Conventions | 8–10 | 10–13 | 16 | 31% | 19–22 | 28–37% |
| **Writing Mechanics Claim:** Students can revise writing products that accurately and convincingly present, describe, and explain ideas for a range of purposes and audiences through Spanish. | Revising and Editing | 5–7 | 6–9 | 16 | 31% | 19–22 | 28–37% |
| **Listening Claim:** Students can comprehend spoken Spanish in a range of contexts. | Listening Comprehension | 12 | 15–17 | 12 | 23% | 15–17 | 22–28% |
| N/A | N/A | N/A | **TOTALS:** | **52** | **100%** | **61–66** | **100%** |

Table 4.A.2 CSA Test Blueprint for Grade Span Six Through Eight

| **Claim/Score Reporting Category** | **Content Category** | **Total Items by Content Category** | **Total Score Points by Content Category** | **Total Items by Claim** | **Percentage of Items by Claim** | **Total Score Points by Claim** | **Percentage of Score Points by Claim** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Reading Claim:** Students can read, analyze, and interpret a variety of texts and genres through Spanish. | Literary | 6–9 | 7–11 | 24 | 46% | 27–35 | 40–58% |
| **Reading Claim:** Students can read, analyze, and interpret a variety of texts and genres through Spanish. | Informational | 6–9 | 7–11 | 24 | 46% | 27–35 | 40–58% |
| **Reading Claim:** Students can read, analyze, and interpret a variety of texts and genres through Spanish. | Vocabulary and Meaning | 8–10 | 10–13 | 24 | 46% | 27–35 | 40–58% |
| **Writing Mechanics Claim:** Students can revise writing products that accurately and convincingly present, describe, and explain ideas for a range of purposes and audiences through Spanish. | Mechanics and Conventions | 7–9 | 8–11 | 16 | 31% | 19–22 | 28–37% |
| **Writing Mechanics Claim:** Students can revise writing products that accurately and convincingly present, describe, and explain ideas for a range of purposes and audiences through Spanish. | Revising and Editing | 7–9 | 8–11 | 16 | 31% | 19–22 | 28–37% |
| **Listening Claim:** Students can comprehend spoken Spanish in a range of contexts. | Listening Comprehension | 12 | 15–17 | 12 | 23% | 15–17 | 22–28% |
| N/A | N/A | N/A | **TOTALS:** | **52** | **100%** | **61–66** | **100%** |

Table 4.A.3 CSA Test Blueprint for Grade Span Nine Through Twelve (High School)

| **Claim/Score Reporting Category** | **Content Category** | **Total Items by Content Category** | **Total Score Points by Content Category** | **Total Items by Claim** | **Percentage of Items by Claim** | **Total Score Points by Claim** | **Percentage of Score Points by Claim** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Reading Claim:** Students can read, analyze, and interpret a variety of texts and genres through Spanish. | Literary | 6–9 | 7–11 | 24 | 46% | 27–35 | 40–58% |
| **Reading Claim:** Students can read, analyze, and interpret a variety of texts and genres through Spanish. | Informational | 6–9 | 7–11 | 24 | 46% | 27–35 | 40–58% |
| **Reading Claim:** Students can read, analyze, and interpret a variety of texts and genres through Spanish. | Vocabulary and Meaning | 8–10 | 10–13 | 24 | 46% | 27–35 | 40–58% |
| **Writing Mechanics Claim:** Students can revise writing products that accurately and convincingly present, describe, and explain ideas for a range of purposes and audiences through Spanish. | Mechanics and Conventions | 7–9 | 8–11 | 16 | 31% | 19–22 | 28–37% |
| **Writing Mechanics Claim:** Students can revise writing products that accurately and convincingly present, describe, and explain ideas for a range of purposes and audiences through Spanish. | Revising and Editing | 7–9 | 8–11 | 16 | 31% | 19–22 | 28–37% |
| **Listening Claim:** Students can comprehend spoken Spanish in a range of contexts. | Listening Comprehension | 12 | 15–17 | 12 | 23% | 15–17 | 22–28% |
| N/A | N/A | N/A | **TOTALS:** | **52** | **100%** | **61–66** | **100%** |

## Test Administration

This chapter details the processes involved in the administration of the 2021–22 California Spanish Assessment (CSA). It also describes the procedures followed by ETS to maintain test security throughout the test administration process.

### Overview

The CSA was administered to students in grades three through eight and high school in 2021–22 in conjunction with the other tests that comprise the California Assessment of Student Performance and Progress (CAASPP) System.

In accordance with the procedures for the computer-based CAASPP, local educational agencies (LEAs) identified test administrators and entered the test administrators as users into the Test Operations Management System (TOMS). ETS provided LEA staff with the appropriate training materials, such as test administration manuals, videos, and webinars, to ensure that the LEA staff and test administrators understood how to administer the computer-based CSA content-area assessments.

The testing window for the 2021–22 administration of the CSA was planned for January 11, 2022, through July 15, 2022. Specific test administration schedules within that window were determined locally pursuant to *California Code of Regulations,* Title 5 (5*CCR*), Section 855(a).

#### In-Person and Remote Testing

In the 2021–22 administration, the California Department of Education (CDE) continued to offer flexibility for testing students to LEAs for the CAASPP (with the exception of the alternate assessments):

1. Students may 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. Students may 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). This option is offered when instruction is received remotely or when there are health issues that do not allow a student to test in person.

### User Roles and Standardization

The test administration procedures were designed so that the tests are administered in a standardized manner. ETS took all necessary measures to ensure the standardization of test administration, as described in this section.

#### Local Educational Agency CAASPP Coordinator

An LEA CAASPP coordinator was designated by the district superintendent or charter school administrator at the beginning of the 2021–22 school year. LEAs include public school districts, California State Board of Education–authorized charter schools, county office of education programs, and direct funded charter schools.

LEA CAASPP coordinators were responsible for ensuring the proper and consistent administration of the CAASPP. In addition to the responsibilities set forth in 5*CCR* Section 857, their responsibilities included

* adding CAASPP test site coordinators and test administrators into TOMS;
* training CAASPP test site coordinators and test administrators regarding the state requirements and CAASPP administration as well as security policies and procedures;
* providing checklists for CAASPP test site coordinators and test administrators to review in preparation for administering the summative assessments;
* overseeing test administration activities;
* reporting test security incidents (including testing irregularities) to the CDE using the online Security and Test Administration Incident Reporting System (STAIRS)/Appeals process;
* filing a report of a testing incident in STAIRS; 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]). A CAASPP test site coordinator’s duties may have included

* adding test administrators into TOMS;
* entering test settings for students;
* creating testing schedules and procedures for a school consistent with state and LEA policies;
* working with technology staff to ensure secure browsers are installed and any technical issues are resolved;
* monitoring testing progress during the testing window and ensuring students take the CSA, 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 were identified by CAASPP test site coordinators as individuals who will administer the CSA.

A test administrator must sign a security affidavit (5 *CCR* Section 859[d]).

A test administrator’s duties may have included

* ensuring the physical conditions of the testing room meet the criteria for a secure test environment;
* administering the CAASPP, including the CSA;
* reporting all test security incidents to the CAASPP 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 CAASPP 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 CSA (CDE, 2022e).

#### Instructions for Test Administrators

##### Test Administrator *Directions for Administration*

The *DFAs,* used by test administrators to administer the CSA to students, are included in the *CAASPP Online Test Administration Manual* (CDE, 2022b). 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. Instructions for the CSA were written in Spanish and were to be read to students in Spanish. 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, 2022b).

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

The *CAASPP Online Test Administration Manual* (CDE, 2022e) 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
* In-person test administration
* Remote test administration
* Test administration directions and scripts for test administrators
* Overview of the student testing application
* Instructions for steps to take before, during, and after testing

Appendices included definitions of common terms and descriptions of different aspects of the test and systems associated with the test.

##### 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, and submit computer-based student test settings.

TOMS modules described in the *TOMS User Guide* included the following (CDE, 2022d):

* **Test Administration Setup—**This module allowed LEAs to determine and calculate dates for the LEA’s 2021–22 administration of the CAASPP, including the CSA.
* **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 computer-based assessments.
* **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 student’s test assignments and test settings.

##### Other System Manuals

Other manuals were created to assist LEA CAASPP coordinators and others with the technological components of the CAASPP System and are listed next.

* ***CAASPP and English Language Proficiency Assessments for California (ELPAC) Technical Specifications and Configuration Guide for Online Testing*—**This manual provided information, tools, and recommended configuration details to help technology staff prepare computers and install the secure browser to be used for the computer-based CAASPP (CDE, 2022c).
* ***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, 2022b).
* ***CAASPP and ELPAC Accessibility Guide*—**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® (software) or a braille embosser (hardware) (CDE, 2022a).

### Local Educational Agency Training

Each year, ETS, in collaboration with the CDE and its Assessment Validity and Outreach contractor, the Sacramento County Office of Education (SCOE), establishes and implements a comprehensive training plan for LEA assessment staff and educators on all aspects of the assessment program. The ETS and SCOE annual training plans specify the audience, topics, frequency, and mode (synchronous or asynchronous) of the training, including such elements as format, participants, and organization.

Knowing that educators were confronted with 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 training, and converting training 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 in one place, on the Upcoming Training Opportunities web page. Archived training was made available on the Past Training Opportunities web page, making it easier for educators to find a training they missed, and providing easier access to recorded training. ETS also employed a new strategy for providing access to training materials. Participants could register to receive a copy of the training materials without registering to attend a live training. Training materials were developed in such a way that educators could consume the information independently by reading through materials.

#### Synchronous and Asynchronous Training

All synchronous training was offered on Zoom, recorded, and made available for on-demand viewing. Zoom provides an opportunity for educators to ask questions and get answers in real time. Training was 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 used various strategies to increase engagement during synchronous 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 used 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.

Working closely with the CDE, ETS and SCOE continued to provide informal support to educators by offering monthly Coffee Sessions. Coffee Sessions included CDE and ETS staff who could answer questions about all aspects of testing. ETS also offered a “workday” for coordinators where support staff were available throughout the day, allowing coordinators to join as needed and get customized support. 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 page on the CAASPP website.

#### Videos and Guides

ETS produced videos on various aspects of administering the CAASPP, including how to perform functions within TOMS, such as setting up a test administration window, adding users, assigning tests to students, and uploading test settings. SCOE produced the accompanying quick reference guides, providing multiple avenues of support for educators administering the assessments.

In addition to the standard administration videos, ETS produced additional videos to support administration. Some videos were geared toward parents and guardians to help them understand the assessment’s purpose or how to read the Student Score Report (SSR). Other videos were intended to help coordinators or other users complete a process, such as administering a practice or training test, starting and stopping a test session, how to monitor student completion, and how to complete second scoring that is required for some of the assessments. This list is a sampling of the available videos intended to capture the major areas of support for various interest holders. The comprehensive suite of training videos can be found on the CAASPP Videos and Quick Reference Guides web page.

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

ETS 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.

ETS also 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. Demonstration videos were also created for the most frequently used non-embedded accessibility resources. These videos were linked within the Individual Student Assessment Accessibility Profile (ISAAP) Tool, increasing access to the demonstration videos. Educators using the ISAAP Tool to determine the student’s needs could view the corresponding demonstration video without having to navigate away from the tool.

A video on how to use the ISAAP Tool was also available to support educators in the process of creating an individual student profile and matching accessibility resources to student needs to ensure a fair and valid testing experience for all students.

For the 2021–22 CAASPP administration, ETS produced a two-part asynchronous training module. Module A, Matching Accessibility Resources to Students’ Needs, 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. Module B, Using Accessibility Resources in Daily Instruction, focused on the importance of removing barriers to student learning and using accessibility resources in daily instruction. Educators could complete the training independently or had the option to attend one of two live sessions held by ETS to extend and deepen the learning experience.

At the California Assessment Conference, SCOE offered two sessions on accessibility. “Leveraging UDL and Accessibility Resources to Improve Teaching and Learning” explored Universal Design for Learning (UDL) principles to help remove barriers to student learning and provided data collection tools to participants. The session on “Introduction to Accessibility and the ISAAP Tool” provided participants with the most up-to-date information regarding accessibility resources and offered a live practical approach to identifying and matching accessibility resources to students using the ISAAP Tool. The conference also included some shared practices sessions focused on accessibility.

#### Feedback for Continuous Improvement Survey

The CAASPP program solicits feedback annually from various interest holder groups, including LEA CAASPP coordinators, CAASPP test site coordinators, test administrators, and test examiners, through the CAASPP and ELPAC Feedback for Continuous Improvement Survey. Feedback was collected via a post-test survey sent to more than 510,000 California educators and through focus groups. Educators provided valuable feedback for potential improvements to the future administration of CAASPP and the ELPAC—one or both—by reporting some lessons they learned in 2021–22.

Improvements made in response to survey results are detailed in [chapter 10](#_Continuous_and_Systematic_1). The CDE and ETS used key recommendations from educators to implement positive changes in the next test administration year.

##### Overview

LEA and CAASPP test site coordinators, as well as test administrators and test examiners, were invited to participate in the survey. The California educators who responded provided specific, actionable insights about their test administration experience. This survey gathered information and data from educators who were part of the administration of CAASPP, the ELPAC, or both programs. Its goal was to highlight successes and identify areas for improvement, both immediate and long term.

Overall, California educators continue to express positive experiences in their preparations for administering CAASPP and the ELPAC.

##### Communication

One example of these efforts is the monthly email communication, which is sent to LEA CAASPP coordinators. During the 2021–22 test administration year, the CDE and ETS worked to streamline communications and provide LEAs with timely and relevant information throughout the year. Monthly CAASPP and ELPAC email communications were sent to the field during the 2021–22 test administration year. More than 70 percent of respondents from the 2021–22 survey found the new monthly email to be helpful.

##### Testing Locations

Although the 2021–22 school year included the transition from distance learning back to in-person learning, educators felt that the resources and training materials provided were useful in preparing them and their students for test administration. Their feedback generally described smooth preparation, training, support, and assessment administration experiences.

With the transition back to in-person learning, testing in person was the primary testing mode for LEAs in 2021–22. However, many LEAs expressed their appreciation for the continued option of remote testing, either for students who were receiving remote instruction or for those with medical parameters that prevented them from attending school in person.

Many of the existing remote testing resources were updated or improved from the previous versions. Most notably, the chapter on remote testing was added to the *CAASPP Online Test Administration Manual,* *Initial ELPAC Online Test Administration Manual,* and the *Summative ELPAC Online Test Administration Manual*. Sixty-nine percent of both LEA CAASPP coordinators and LEA ELPAC coordinators indicated that they found it helpful.

### Accessibility Resources

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 in the CSA, as well as the procedures to identify and assign students with accommodations and designated supports. Finally, the number of students who were assigned accessibility resources was reported on the basis of available data.

The 2021–22 CSA offered commonly used accessibility resources available through the CAASPP computer-based testing platform, where applicable for the tested construct.

#### Accessibility Resource Categories

The purpose of universal tools, designated supports, and accommodations in testing is to provide *all* students with the opportunity to demonstrate what they know and what they are able to do. Universal tools, designated supports, and accommodations minimize or remove barriers that could otherwise prevent students from demonstrating their knowledge, skills, and achievement in a specific content area.

The CDE’s *California Assessment Accessibility Resources Matrix* (Accessibility Matrix) (CDE, 2021) is intended for school-level personnel and individualized education program (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 2021–22 CSA administration.

###### Embedded

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

* Breaks
* Digital notepad
* Expandable items
* Expandable passages
* Highlighter
* Keyboard navigation
* Line reader
* Mark for review
* Spanish glossary (for specific items)
* Strikethrough
* Zoom (in or out)

###### Non-Embedded

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

* Breaks
* Scratch paper

##### Designated Supports

Designated supports were available to all students through the test settings in TOMS. The designated supports each fell into one of two categories: embedded and non-embedded. Embedded designated supports were provided through the Student Testing Interface (through the CAASPP secure browser).

The designated supports in the following subsections were available in the 2021–22 CSA administration.

###### Embedded

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

* Color contrast
* Masking
* Mouse pointer (size and color)
* Permissive mode
* Print (font) size
* Streamline
* Text-to-speech (items)
* Turn off any universal tool(s)

###### Non-Embedded

The following non-embedded designated supports were available to students testing in the secure browser:

* Amplification
* Color contrast
* Color overlay
* Magnification
* Medical supports
* Noise buffers
* Read aloud (items)
* Scribe (nonwriting items)
* Separate setting (special lighting or acoustics, adaptive furniture, time of day)
* Simplified test directions

##### Accommodations

Accommodations are changes in procedures or materials that increased equitable access during CAASPP administration. 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 2021–22 CSA administration.

###### Embedded

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

* Audio transcript
* Braille (embossed and refreshable)
* Closed-captioning
* Text-to-speech (reading passages)

###### Non-Embedded

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

* Alternate response options
* Print-on-demand
* Read aloud (reading passages)

##### Unlisted Resources

An unlisted resource is an instructional support a student regularly uses in daily instruction, assessment, or both, and has not been previously identified as a universal tool, designated support, or accommodation. The Accessibility Matrix included an inventory of unlisted resources that were already identified and were preapproved (CDE, 2021). During the 2021–22 CAASPP administration, an LEA CAASPP coordinator or a CAASPP test site coordinator would use TOMS to submit a request for use of an unlisted resource. A preidentified, preapproved unlisted resource request was automatically approved. A request for an unlisted resource that was not preidentified was sent to the CDE for review and adjudication.

Unlisted resources are non-embedded resources that are made available if specified in the eligible student’s IEP or Section 504 plan and only upon approval by the CDE. Unlisted resources that changed the construct of an assessment and were approved were flagged as causing a change in construct. Test results for a student using an unlisted resource that was approved but that changed the construct of what was being tested were considered invalid for reporting purposes. The lowest obtainable scale score (LOSS) would be assigned to the CSA with the unlisted resource that changes the construct, the student’s score status would remain valid, and the student’s scale score would be reported but appear on the 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.

Use of an English dictionary is the only preidentified unlisted resource applicable to the CSA.

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 LOSS was reported for the affected domain when administrations included unlisted resources that changed the construct of that assessment.

#### Identification and Selection

All eligible students enrolled in a California public school participate in the CAASPP System, including students with disabilities and English learner students. The Smarter Balanced Assessment Consortium’s *Usability, Accessibility, and Accommodations Guidelines* (Smarter Balanced, 2021) and the CDE Accessibility Matrix (CDE, 2021) 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.[[3]](#footnote-4) The CSA follows the Smarter Balanced recommendations for use (Smarter Balanced, 2018).

The *Guidelines* apply to all participating students and promote an individualized approach to the implementation of assessment practices. Another web page, the Smarter Balanced Accessibility Strategies web page on the Tools for Teachers website (Smarter Balanced, 2022), connects the assessment resources described in the *Guidelines* with associated classroom practices.

The full list of the universal tools, designated supports, and accommodations used in CAASPP computer-based assessments, including the CSA, is documented in the Accessibility Matrix. Most embedded and non-embedded universal tools, designated supports, and accommodations listed in parts 1, 2, and 3 of the Accessibility Matrix are available for the CSA 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. Another manual, the *Smarter Balanced Usability, Accessibility, and Accommodations Implementation Guide* (Smarter Balanced, 2014),provides suggestions for implementation of these resources.

Test administrators and test examiners are given the opportunity to administer the CAASPP practice and training tests so that students have the opportunity to familiarize themselves with a designated support or accommodation prior to testing.

#### 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 through the student’s profile in TOMS or in a batch upload for multiple students. When the batch upload process was used, settings were uploaded into TOMS using a spreadsheet with data that had either been entered into a template downloaded from TOMS; or created by selecting and entering information into the web-based ISAAP Tool. The ISAAP Tool could be used by LEAs in conjunction with the *Guidelines* and the *2021–22* CAASPP and ELPAC Accessibility Guide (CDE, 2022a), 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_1) in [*Chapter 1: Introduction*](#_Introduction) for more details regarding the TDS.

Once a student’s IEP or Section 504 plan team decided which accessibility resource(s) the student should use, LEA CAASPP coordinators and 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 a student’s accessibility resource(s) could be assigned:

1. Using the ISAAP Tool to identify the accessibility resource(s) and then uploading the spreadsheet it creates into TOMS (This process is discussed in more detail in subsection [*5.4.2 Identification and Selection*](#_Identification_and_Selection).)
2. Using the Online Student Test Settings template to enter students’ assignments and then uploading the spreadsheet into TOMS
3. Entering assignments for each student individually in TOMS

If a student’s IEP or Section 504 plan team identified and designated a resource not identified in the CDE Accessibility Matrix, the LEA CAASPP coordinator or 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.

[Appendix 5.A](#_Appendix_5.A:_Accessibility) provides information on the number of students who were assigned accommodations and designated supports.

#### Delivery of Embedded and Non-Embedded Resources to Students

Universal tools, designated supports, and accommodations can be delivered as either embedded or non-embedded resources. Embedded resources are digitally delivered features or settings available as part of the technology platform for CSA testing. Examples of embedded resources include the expandable items, color contrast, and masking.

Non-embedded resources are available, when provided by the LEA, for both computer-based assessments and paper–pencil tests. These resources are not part of the technology platform for the computer-administered CSA. Examples of non-embedded resources include magnification, noise buffers, and the use of a scribe.

Refer to subsection [*5.4.1 Accessibility Resource Categories*](#_Accessibility_Resource_Categories_1) for a detailed description of the accessibility resources available to students taking the CSA.

#### 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) was used by a student as the student progressed through the test.

[Appendix 5.A](#_Appendix_5.A:_Accessibility) reports the number of students who were assigned to a certain accommodation or designated support at school. Table 5.2 reports the number of students who, based on the availability of data, actually used this accommodation or designated support at least once. The number of students in [appendix 5.A](#_Appendix_5.A:_Accessibility) is higher than the number of students in table 5.2 because a student who was assigned to an accessibility resource may not have used it during testing. Table 5.2 reports the available data only; the first row of a grade level or the high school grade band (i.e., grades nine through twelve), the *All* row, combines data across the resources listed in this table and counts each student once only. However, because the TDS does not capture the usage of all embedded resources and cannot capture the usage of any non-embedded resources, this table reports only on a limited subset of the embedded resources. (Embedded accessibility resources are those that are part of the computer-based TDS, whereas non-embedded accessibility resources are provided outside of that system.)

Types of accommodations and designated supports—labeled “ACC” and “DS” in the *Resource Type* column—included in table 5.2 are as follows:

1. **Audio Transcript:** This resource allows students to view a transcript of the audio content for the current test page. This is useful for students with visual impairment who are accustomed to accessing information presented via audio in the form of braille.
2. **Text-to-Speech:** Text is read aloud to the student via embedded text-to-speech technology.
3. **Print-on-Demand:** Paper copies of passages and stimuli, items, or all of these are printed for students.
4. **Masking:** This resource involves blocking off content that is not of immediate need or that may be distracting to the student.

Table 5.2 Summary of Accommodations and Designated Supports Used by Students, by Grade Level or Grade Band

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grade Level or Grade Band | Accessibility Resource | Resource Type | Students Assigned | Students Used | Percentage Used |
| All | Any Tracked Resource | All | 3,717 | 1,010 | 27.17 |
| All | Embedded Audio Transcript | ACC | 4 | 0 | 0.00 |
| All | Embedded Text-to-Speech Passages | ACC | 234 | 139 | 59.40 |
| All | Non-Embedded Print-on-Demand | ACC | 5 | 2 | 40.00 |
| All | Embedded Masking | DS | 625 | 25 | 4.00 |
| All | Embedded Text-to-Speech Items | DS | 3,551 | 970 | 27.32 |
| Grade 3 | Any Tracked Resource | All | 1,312 | 456 | 34.76 |
| Grade 3 | Embedded Audio Transcript | ACC | 1 | 0 | 0.00 |
| Grade 3 | Embedded Text-to-Speech Passages | ACC | 66 | 35 | 53.03 |
| Grade 3 | Non-Embedded Print-on-Demand | ACC | 1 | 0 | 0.00 |
| Grade 3 | Embedded Masking | DS | 125 | 3 | 2.40 |
| Grade 3 | Embedded Text-to-Speech Items | DS | 1,274 | 451 | 35.40 |
| Grade 4 | Any Tracked Resource | All | 985 | 246 | 24.97 |
| Grade 4 | Embedded Audio Transcript | ACC | 1 | 0 | 0.00 |
| Grade 4 | Embedded Text-to-Speech Passages | ACC | 54 | 35 | 64.81 |
| Grade 4 | Non-Embedded Print-on-Demand | ACC | 3 | 2 | 66.67 |
| Grade 4 | Embedded Masking | DS | 157 | 6 | 3.82 |
| Grade 4 | Embedded Text-to-Speech Items | DS | 945 | 235 | 24.87 |
| Grade 5 | Any Tracked Resource | All | 692 | 190 | 27.46 |
| Grade 5 | Embedded Audio Transcript | ACC | 1 | 0 | 0.00 |
| Grade 5 | Embedded Text-to-Speech Passages | ACC | 66 | 46 | 69.70 |
| Grade 5 | Non-Embedded Print-on-Demand | ACC | 1 | 0 | 0.00 |
| Grade 5 | Embedded Masking | DS | 103 | 6 | 5.83 |
| Grade 5 | Embedded Text-to-Speech Items | DS | 662 | 180 | 27.19 |

Table 5.2 *(continuation one)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grade Level or Grade Band | Accessibility Resource | Resource Type | Students Assigned | Students Used | Percentage Used |
| Grade 6 | Any Tracked Resource | All | 493 | 91 | 18.46 |
| Grade 6 | Embedded Audio Transcript | ACC | 0 | 0 | N/A |
| Grade 6 | Embedded Text-to-Speech Passages | ACC | 31 | 20 | 64.52 |
| Grade 6 | Non-Embedded Print-on-Demand | ACC | 0 | 0 | N/A |
| Grade 6 | Embedded Masking | DS | 90 | 5 | 5.56 |
| Grade 6 | Embedded Text-to-Speech Items | DS | 472 | 82 | 17.37 |
| Grade 7 | Any Tracked Resource | All | 120 | 18 | 15.00 |
| Grade 7 | Embedded Audio Transcript | ACC | 1 | 0 | 0.00 |
| Grade 7 | Embedded Text-to-Speech Passages | ACC | 11 | 2 | 18.18 |
| Grade 7 | Non-Embedded Print-on-Demand | ACC | 0 | 0 | N/A |
| Grade 7 | Embedded Masking | DS | 75 | 4 | 5.33 |
| Grade 7 | Embedded Text-to-Speech Items | DS | 105 | 15 | 14.29 |
| Grade 8 | Any Tracked Resource | All | 96 | 8 | 8.33 |
| Grade 8 | Embedded Audio Transcript | ACC | 0 | 0 | N/A |
| Grade 8 | Embedded Text-to-Speech Passages | ACC | 6 | 1 | 16.67 |
| Grade 8 | Non-Embedded Print-on-Demand | ACC | 0 | 0 | N/A |
| Grade 8 | Embedded Masking | DS | 68 | 1 | 1.47 |
| Grade 8 | Embedded Text-to-Speech Items | DS | 81 | 6 | 7.41 |
| Grade 9 | Any Tracked Resource | All | 1 | 1 | 100.00 |
| Grade 9 | Embedded Audio Transcript | ACC | 0 | 0 | N/A |
| Grade 9 | Embedded Text-to-Speech Passages | ACC | 0 | 0 | N/A |
| Grade 9 | Non-Embedded Print-on-Demand | ACC | 0 | 0 | N/A |
| Grade 9 | Embedded Masking | DS | 0 | 0 | N/A |
| Grade 9 | Embedded Text-to-Speech Items | DS | 1 | 1 | 100.00 |

Table 5.2 *(continuation two)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grade Level or Grade Band | Accessibility Resource | Resource Type | Students Assigned | Students Used | Percentage Used |
| Grade 10 | Any Tracked Resource | All | 0 | 0 | N/A |
| Grade 10 | Embedded Audio Transcript | ACC | 0 | 0 | N/A |
| Grade 10 | Embedded Text-to-Speech Passages | ACC | 0 | 0 | N/A |
| Grade 10 | Non-Embedded Print-on-Demand | ACC | 0 | 0 | N/A |
| Grade 10 | Embedded Masking | DS | 0 | 0 | N/A |
| Grade 10 | Embedded Text-to-Speech Items | DS | 0 | 0 | N/A |
| Grade 11 | Any Tracked Resource | All | 17 | 0 | 0.00 |
| Grade 11 | Embedded Audio Transcript | ACC | 0 | 0 | N/A |
| Grade 11 | Embedded Text-to-Speech Passages | ACC | 0 | 0 | N/A |
| Grade 11 | Non-Embedded Print-on-Demand | ACC | 0 | 0 | N/A |
| Grade 11 | Embedded Masking | DS | 6 | 0 | 0.00 |
| Grade 11 | Embedded Text-to-Speech Items | DS | 11 | 0 | 0.00 |
| Grade 12 | Any Tracked Resource | All | 1 | 0 | 0.00 |
| Grade 12 | Embedded Audio Transcript | ACC | 0 | 0 | N/A |
| Grade 12 | Embedded Text-to-Speech Passages | ACC | 0 | 0 | N/A |
| Grade 12 | Non-Embedded Print-on-Demand | ACC | 0 | 0 | N/A |
| Grade 12 | Embedded Masking | DS | 1 | 0 | 0.00 |
| Grade 12 | Embedded Text-to-Speech Items | DS | 0 | 0 | N/A |
| High school | Any Tracked Resource | All | 19 | 1 | 5.26 |
| High school | Embedded Audio Transcript | ACC | 0 | 0 | N/A |
| High school | Embedded Text-to-Speech Passages | ACC | 0 | 0 | N/A |
| High school | Non-Embedded Print-on-Demand | ACC | 0 | 0 | N/A |
| High school | Embedded Masking | DS | 7 | 0 | 0.00 |
| High school | Embedded Text-to-Speech Items | DS | 12 | 1 | 8.33 |

### Practice and Training Tests

Practice and training tests are available publicly to LEA staff, students, parent/guardians, and any other individual for the CSA. These tests simulate the experience of the computer-based CSA to allow anyone to experience the test. For the 2021–22 school year, accommodated versions of CSA practice and training tests were developed to include all accessibility resources—including braille, closed-captioning, text-to-speech, and audio transcripts—available on the assessment.

Students can access practice and training tests using a web browser. They allow students and administrators to familiarize themselves with the user interface and components of the TDS and help maintain the standardization of test administration. Practice and training tests are available through the Practice and Training Test website linked on the Online Practice and Training Tests Portal web page on the CAASPP website.

The practice tests, offered at each grade level or grade span, were released to prepare students for the CSA. These tests more closely simulate the CSA’s length and complexity and align with the CSA blueprint.

The grade-level-specific training tests can be taken by students in all tested grade levels or grade band. Many unique item types available on the operational test are covered in the training tests. The scoring guides for the practice and training tests are available on the Online Practice Test Scoring Guides and *Directions for Administration (DFAs)* web page on the CAASPP website.

### Test Security and Confidentiality

For the operational CSA, 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 configurations of secure servers), were kept secure. ETS had systems in place that maintained tight security for test items and test results, as well as for student data. To ensure security for all tests that ETS develops or handles, ETS maintains an Office of Testing Integrity (OTI), which is described in the next subsection.

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 *CSA 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 testing programs managed by ETS. This division resides in the ETS legal department. The Office of Professional Standards Compliance at ETS publishes and maintains the *ETS Standards for Quality and Fairness* (2014), which supports 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 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 its pursuit of enforcing secure testing practices, the OTI strives to safeguard the various processes involved in a test development and administration cycle. For the CSA, 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 and student data—before, during, and after each test administration. The LEA CAASPP coordinator is responsible for keeping all electronic test materials secure, keeping student information confidential, and making sure the CAASPP test site coordinators and test administrators are 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.

The test administrator is responsible for reporting testing incidents to the CAASPP test site coordinator and securely destroying printed and digital media for items and passages generated by the print-on-demand feature of the TDS (CDE, 2022e).

The following measures ensured the security of the CAASPP:

* LEA CAASPP coordinators and CAASPP 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 can grant the coordinators access to TOMS (5 *CCR* Section 859[d]).
* Anyone having access to the testing materials must have electronically signed and submitted a “Test Security Affidavit for Test Examiners, Test Administrators, Proctors, Translators, Scribes, and Any Other Person Having Access to CAASPP Tests” form in TOMS before receiving access to any testing materials (5 *CCR* Section 859[c]).
* Anyone having access to the testing materials but not having access to TOMS must have signed the CAASPP *Test Security Affidavit for Non-TOMS Users*, which was available as a web-based form, before receiving access to any testing materials.

In addition, it was the responsibility of every participant in the CAASPP System to report immediately any violation or suspected violation of test security or confidentiality. The test administrator reported to the 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 occurred. ETS performed site visits and testing procedure audits at randomly selected LEAs and test sites throughout California during the test administration of the ELPAC and the CAASPP operational assessments. Audits were performed before, during, and after test administrations to observe adherence to published procedures regarding the handling of testing materials and test administration guidelines.

To provide this service, ETS used its OTI and subcontractor staff as auditors. All auditors had a minimum of a high school diploma, a valid driver’s license, and experience in security auditing or a related field. All had passed a background check conducted by the subcontracted vendor as part of the employment process.

ETS provided a final summary report of audit findings to the CDE at the end of the test administration. In addition, the OTI reported findings and recommendations to ETS program management on a weekly basis as audits were completed. ETS program management reported a summary of these findings to the CDE after a site visit. The OTI also provided individual audit reports directly to the LEA at the completion of the testing year.

#### 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 2021–22 CSA, ETS entered information about the deliverable into a web form on a SharePoint website when a file was posted. A CDE staff member monitored this log throughout the day for updates to the status of deliverables and downloaded and deleted the file from the SFTP server when its status showed that it had been posted.

#### Data Management in the Secure Database

ETS 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 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, 2022d).

#### 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. The statistical analysis staff store the files on secure servers. All staff members involved with the data adhere to the ETS Code of Ethics and the ETS Information Protection Policies to prevent any unauthorized access to data.

#### Student Confidentiality

To meet the requirements of the Every Student Succeeds Act, as well as state requirements, LEAs must collect demographic data about students’ ethnicity, disabilities, parent/guardian education, and so forth during the school year. ETS takes every precaution to prevent any of this information from becoming public or being used for anything other than for testing and score-reporting purposes. These procedures are applied to all documents in which student demographic data appears, such as technical reports.

#### Student Test Results

##### Types of Results

The following deliverables are produced for reporting of the CSA:

* Individual student results for computer-based assessments in the California Educator Reporting System
* Individual SSRs (electronic)
* Internet reports—available on the CDE Test Results for California’s Assessments website—aggregated by content area and state, county, LEA, or test site

##### Security of Results Files

ETS takes measures to protect files and reports that show students’ scores and reporting levels. ETS is committed to safeguarding all secure information in its possession from unauthorized access, disclosure, modification, or destruction. ETS has strict information security policies in place to protect the confidentiality of both student and client data. 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. 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, 2022b). 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 the LEA Outreach Administrator with the opportunity to interact and communicate regarding the STAIRS process (CDE, 2022b).

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

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

Appeals requests were reviewed by the CDE or an ETS LEA Outreach Administrator. When a request to submit an Appeal was approved, the coordinator received a system-generated email with the Appeal type that was approved (CDE, 2022b).

Types of Appeals available during the 2021–22 CAASPP administration are described in table 5.3.

Table 5.3 Types of Appeals

|  |  |
| --- | --- |
| **Type of Appeal** | **Description** |
| Reset | Resetting a student’s assessment removed that assessment from the system and enabled the student to start a new assessment from the beginning. |
| Invalidate | Invalidated assessments were scored, and scores were provided on the SSR with a note that an irregularity occurred. The student(s) was counted as participating in the calculation of the school’s participation rate for accountability purposes. |
| Re-open | Reopening an assessment allowed a student to access an assessment that had already been submitted or had expired. |
| Restore | Restoring an assessment returned an assessment from the Reset status to its prior status. This action could be performed only on 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.  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 be reported to the LEA CAASPP coordinator and CAASPP test site coordinator immediately. The coordinator must report the incident within 24 hours, using the STAIRS/Appeals process in TOMS.

##### Irregularity

A testing irregularity is an unusual circumstance that impacts an individual or a group of students who are testing and may potentially affect student performance on the 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 must be reported to the LEA CAASPP coordinator and CAASPP test site coordinator immediately. The coordinator must report the irregularity within 24 hours, using the online STAIRS/Appeals process in TOMS.

##### Breach

A testing breach is an event that poses a threat to the validity of the test. Breaches require immediate attention; a breach that was due to social media exposure on the part of a student or adult or due to media coverage of an administration was to be escalated to CalTAC via telephone. Following the call, the CAASPP test site coordinator or LEA CAASPP coordinator must report the incident using the online STAIRS/Appeals process in TOMS within 24 hours. All other breaches were to be entered into STAIRS directly.

Examples may include such situations as a release of secure materials or a security or system risk. These circumstances have external implications for the CDE and may result in a decision to remove the test item(s) from the available secure item bank.

#### Appeals

For test security incidents reported in STAIRS that resulted in a need to reset, reopen, invalidate, or restore individual computer-based student assessments, the request had to be approved by the CDE. Requests to reset and reopen assessments were processed by an LEA Outreach Administrator.

In most instances, an Appeal was submitted to address a test security breach or irregularity. The LEA CAASPP coordinator or CAASPP test site coordinator submitted Appeals in TOMS. All submitted Appeals were available for retrieval and reviewed by LEA and site coordinators within a given organization. However, the view of Appeals was restricted according to the user role as established in TOMS (CDE, 2022a). 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, 2022a). Types of Appeals available during the 2021–22 CAASPP test administration are described in table 5.3.

Table 5.4 presents the number and types of incidents submitted in STAIRS in the 2021–‍22 test administration for the CSA as well as the number of individual Statewide Student Identifiers (SSIDs) submitted and approved. The number in the *Appeals SSIDs Approved* column is the number of accepted cases that resulted in an Appeal, which may differ from the number in the *Number of Incidents* column because of incorrect entry or other factors.

Table 5.4 Number and Types of Incidents Submitted in STAIRS in the 2021–22 Test Administration

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Description** | **Appeal Type** | **Number of Incidents** | **Total Number of SSIDs Submitted** | **Appeals SSIDs Approved** |
| Accessibility Issue | Reset | 2 | 2 | 2 |
| Administered Incorrect Assessment | Reset | 0 | 0 | 0 |
| Administration Error | No Appeal | 0 | 0 | 0 |
| Data Entry Issue | Reset or Re-open | 0 | 0 | 0 |
| Expired or Accidentally Submitted Test | Re-open | 3 | 23 | 23 |
| Exposing Secure Materials | Invalidate or No Appeal | 0 | 0 | 0 |
| Incorrect SSID Used | Reset or No Appeal | 0 | 0 | 0 |
| Restore from Reset | Restore | 0 | 0 | 0 |
| Student Cheating or Accessing Unauthorized Devices | Invalidate | 1 | 1 | 1 |
| Student Disruption | No Appeal | 0 | 0 | 0 |
| Technical Issues | Grace Period Extension or No Appeal | 4 | 1 | 0 |
| Validity Issue | Invalidate or Reset | 1 | 1 | 1 |
| **Total:** | **N/A** | **11** | **28** | **27** |

Table 5.5 presents the number of Appeals approved and rejected in STAIRS in the 2021–22 administration of the CSA, for all grade levels combined.

Table 5.5 Number of Appeals Approved in STAIRS in the 2021–22 Test Administration

|  |  |  |
| --- | --- | --- |
| **Appeal Type** | **Number of Appeals Approved for the CSA** | **Number of Appeals Rejected for the CSA** |
| Reset | 2 | 0 |
| Re-open | 23 | 0 |
| Invalidate | 2 | 0 |
| Grace Period Extension | 0 | 1 |
| Restore | 0 | 0 |
| **Total:** | **27** | **1** |

### References

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

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

California Department of Education. (2021). *California assessment accessibility resources matrix*. Sacramento, CA: California Department of Education.

California Department of Education. (2022a). *CAASPP and ELPAC accessibility guide*. Sacramento, CA: California Department of Education.

California Department of Education. (2022b). *CAASPP and ELPAC security incidents and Appeals procedure guide.* Sacramento, CA: California Department of Education.

California Department of Education. (2022c). *CAASPP and ELPAC technical specifications and configuration guide for online testing*. Sacramento, CA: California Department of Education.

California Department of Education. (2022d). *CAASPP and ELPAC Test Operations Management System user guide*. Sacramento, CA: California Department of Education.

California Department of Education. (2022e). *CAASPP* *online test administration manual.* 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. (2021). *Smarter Balanced Assessment Consortium: Usability, accessibility, and accommodations guidelines.* Los Angeles: Smarter Balanced Assessment Consortium.

Smarter Balanced Assessment Consortium. (2022). *Accessibility strategies* (web page). Los Angeles: Smarter Balanced Assessment Consortium.

### Appendix 5.A: Accessibility Resource Assignment

Table 5.A.1 Special Services Summary, Grades Three Through Five

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Accessibility Resource** | **Grade 3 Number** | **Grade 3 Pct. of Total** | **Grade 4 Number** | **Grade 4 Pct. of Total** | **Grade 5 Number** | **Grade 5 Pct. of Total** |
| Embedded Accommodation—Audio Transcript | 1 | 0.01 | 1 | 0.01 | 1 | 0.01 |
| Embedded Accommodation—Braille | 0 | 0.00 | 1 | 0.01 | 0 | 0.00 |
| Embedded Accommodation—Closed Captioning | 4 | 0.04 | 4 | 0.04 | 5 | 0.06 |
| Embedded Accommodation—Text-to-Speech | 66 | 0.59 | 54 | 0.53 | 66 | 0.78 |
| Non-Embedded Accommodation—Alternate Response Options | 1 | 0.01 | 0 | 0.00 | 0 | 0.00 |
| Non-Embedded Accommodation—Print-on-Demand | 1 | 0.01 | 3 | 0.03 | 1 | 0.01 |
| Non-Embedded Accommodation—Read Aloud (for Passages) | 48 | 0.43 | 40 | 0.40 | 42 | 0.50 |
| Non-Embedded Accommodation—Scribe (for Writing) | 17 | 0.15 | 14 | 0.14 | 11 | 0.13 |
| Embedded Designated Support—Color Contrast | 1 | 0.01 | 8 | 0.08 | 21 | 0.25 |
| Embedded Designated Support—Masking | 125 | 1.13 | 157 | 1.55 | 103 | 1.22 |
| Embedded Designated Support—Mouse Pointer | 153 | 1.38 | 108 | 1.07 | 104 | 1.23 |
| Embedded Designated Support—Permissive Mode | 7 | 0.06 | 7 | 0.07 | 4 | 0.05 |
| Embedded Designated Support—Print Size | 23 | 0.21 | 67 | 0.66 | 27 | 0.32 |
| Embedded Designated Support—Streamlining | 63 | 0.57 | 82 | 0.81 | 62 | 0.74 |
| Embedded Designated Support—Text-to-Speech | 1,274 | 11.47 | 945 | 9.34 | 662 | 7.85 |
| Embedded Designated Support—Turn off Universal Tools | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |

Table 5.A.1 *(continuation)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Accessibility Resource** | **Grade 3 Number** | **Grade 3 Pct. of Total** | **Grade 4 Number** | **Grade 4 Pct. of Total** | **Grade 5 Number** | **Grade 5 Pct. of Total** |
| Non-Embedded Designated Support—Amplification | 6 | 0.05 | 18 | 0.18 | 73 | 0.87 |
| Non-Embedded Designated Support—Color Contrast | 1 | 0.01 | 0 | 0.00 | 1 | 0.01 |
| Non-Embedded Designated Support—Color Overlay | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Non-Embedded Designated Support—Magnification | 23 | 0.21 | 27 | 0.27 | 21 | 0.25 |
| Non-Embedded Designated Support—Medical Device | 1 | 0.01 | 0 | 0.00 | 1 | 0.01 |
| Non-Embedded Designated Support—Noise Buffers | 54 | 0.49 | 67 | 0.66 | 138 | 1.64 |
| Non-Embedded Designated Support—Read Aloud | 240 | 2.16 | 111 | 1.10 | 90 | 1.07 |
| Non-Embedded Designated Support—Scribe | 18 | 0.16 | 16 | 0.16 | 17 | 0.20 |
| Non-Embedded Designated Support—Separate Setting | 208 | 1.87 | 178 | 1.76 | 161 | 1.91 |
| Other—Unlisted Resources | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Other—Designated support or accommodation is in IEP | 263 | 2.37 | 231 | 2.28 | 204 | 2.42 |
| Other—Designated support or accommodation is in Section 504 plan | 19 | 0.17 | 25 | 0.25 | 13 | 0.15 |

Table 5.A.2 Special Services Summary, Grades Six Through Eight

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Accessibility Resource** | **Grade 6 Number** | **Grade 6 Pct. of Total** | **Grade 7 Number** | **Grade 7 Pct. of Total** | **Grade 8 Number** | **Grade 8 Pct. of Total** |
| Embedded Accommodation—Audio Transcript | 0 | 0.00 | 1 | 0.03 | 0 | 0.00 |
| Embedded Accommodation—Braille | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Embedded Accommodation—Closed Captioning | 3 | 0.05 | 3 | 0.08 | 1 | 0.03 |
| Embedded Accommodation—Text-to-Speech | 31 | 0.53 | 11 | 0.29 | 6 | 0.19 |
| Non-Embedded Accommodation—Alternate Response Options | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Non-Embedded Accommodation—Print-on-Demand | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Non-Embedded Accommodation—Read Aloud (for Passages) | 18 | 0.31 | 9 | 0.23 | 6 | 0.19 |
| Non-Embedded Accommodation—Scribe (for Writing) | 1 | 0.02 | 0 | 0.00 | 1 | 0.03 |
| Embedded Designated Support—Color Contrast | 40 | 0.68 | 1 | 0.03 | 0 | 0.00 |
| Embedded Designated Support—Masking | 90 | 1.53 | 75 | 1.96 | 68 | 2.10 |
| Embedded Designated Support—Mouse Pointer | 128 | 2.17 | 62 | 1.62 | 61 | 1.88 |
| Embedded Designated Support—Permissive Mode | 6 | 0.10 | 7 | 0.18 | 57 | 1.76 |
| Embedded Designated Support—Print Size | 5 | 0.08 | 0 | 0.00 | 1 | 0.03 |
| Embedded Designated Support—Streamlining | 17 | 0.29 | 71 | 1.85 | 68 | 2.10 |
| Embedded Designated Support—Text-to-Speech | 472 | 8.01 | 105 | 2.74 | 81 | 2.50 |
| Embedded Designated Support—Turn off Universal Tools | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Non-Embedded Designated Support—Amplification | 2 | 0.03 | 1 | 0.03 | 1 | 0.03 |
| Non-Embedded Designated Support—Color Contrast | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Non-Embedded Designated Support—Color Overlay | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Non-Embedded Designated Support—Magnification | 1 | 0.02 | 1 | 0.03 | 1 | 0.03 |
| Non-Embedded Designated Support—Medical Device | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Non-Embedded Designated Support—Noise Buffers | 31 | 0.53 | 66 | 1.72 | 62 | 1.91 |
| Non-Embedded Designated Support—Read Aloud | 34 | 0.58 | 74 | 1.93 | 66 | 2.04 |
| Non-Embedded Designated Support—Scribe | 3 | 0.05 | 0 | 0.00 | 1 | 0.03 |
| Non-Embedded Designated Support—Separate Setting | 111 | 1.88 | 53 | 1.38 | 41 | 1.27 |

Table 5.A.2 *(continuation)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Accessibility Resource** | **Grade 6 Number** | **Grade 6 Pct. of Total** | **Grade 7 Number** | **Grade 7 Pct. of Total** | **Grade 8 Number** | **Grade 8 Pct. of Total** |
| Other—Unlisted Resources | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Other—Designated support or accommodation is in IEP | 143 | 2.43 | 73 | 1.90 | 47 | 1.45 |
| Other—Designated support or accommodation is in Section 504 plan | 17 | 0.29 | 5 | 0.13 | 4 | 0.12 |

Table 5.A.3 Special Services Summary, High School

|  |  |  |
| --- | --- | --- |
| **Accessibility Resource** | **High School Number** | **High School Pct. of Total** |
| Embedded Accommodation—Audio Transcript | 0 | 0.00 |
| Embedded Accommodation—Braille | 0 | 0.00 |
| Embedded Accommodation—Closed Captioning | 2 | 0.06 |
| Embedded Accommodation—Text-to-Speech | 0 | 0.00 |
| Non-Embedded Accommodation—Alternate Response Options | 0 | 0.00 |
| Non-Embedded Accommodation—Print-on-Demand | 0 | 0.00 |
| Non-Embedded Accommodation—Read Aloud (for Passages) | 1 | 0.03 |
| Non-Embedded Accommodation—Scribe (for Writing) | 0 | 0.00 |
| Embedded Designated Support—Color Contrast | 0 | 0.00 |
| Embedded Designated Support—Masking | 7 | 0.20 |
| Embedded Designated Support—Mouse Pointer | 0 | 0.00 |
| Embedded Designated Support—Permissive Mode | 6 | 0.17 |
| Embedded Designated Support—Print Size | 0 | 0.00 |
| Embedded Designated Support—Streamlining | 2 | 0.06 |
| Embedded Designated Support—Text-to-Speech | 12 | 0.34 |
| Embedded Designated Support—Turn off Universal Tools | 0 | 0.00 |
| Non-Embedded Designated Support—Amplification | 0 | 0.00 |
| Non-Embedded Designated Support—Color Contrast | 0 | 0.00 |
| Non-Embedded Designated Support—Color Overlay | 0 | 0.00 |
| Non-Embedded Designated Support—Magnification | 0 | 0.00 |
| Non-Embedded Designated Support—Medical Device | 0 | 0.00 |
| Non-Embedded Designated Support—Noise Buffers | 1 | 0.03 |
| Non-Embedded Designated Support—Read Aloud | 1 | 0.03 |
| Non-Embedded Designated Support—Scribe | 0 | 0.00 |
| Non-Embedded Designated Support—Separate Setting | 12 | 0.34 |

Table 5.A.3 *(continuation)*

|  |  |  |
| --- | --- | --- |
| **Accessibility Resource** | **High School Number** | **High School Pct. of Total** |
| Other—Unlisted Resources | 0 | 0.00 |
| Other—Designated support or accommodation is in IEP | 15 | 0.43 |
| Other—Designated support or accommodation is in Section 504 plan | 0 | 0.00 |

## Standard Setting

### Description

Standard setting, which also is referred to as reporting level setting, refers to a class of methodologies by which one or more thresholds are used to determine achievement levels. The California Department of Education (CDE) set three reporting ranges—range 1, range 2, and range 3—with two threshold cuts for each grade level and content area.

The CDE and ETS implemented an extensive level–setting process involving software development, item mapping, review panels, committees, workshops, and extensive validity research to set the final thresholds and achievement level descriptors. For detailed information regarding this process, refer to the *Standard Setting Technical Report for the California Spanish Assessment* (CDE, 2019).

### Reference

California Department of Education. (2019). *Standard setting technical report for the California Spanish Assessment.* California Department of Education website.

## Scoring and Reporting

To determine individual students’ scores for the California Spanish Assessment (CSA), student item responses were scored, and individual student scores were calculated on the basis of the item responses. In addition, student test scores were aggregated to produce information for schools and local educational agencies (LEAs).

This chapter describes how various types of student responses were scored, as well as the various types of scores and score reports that were produced at the end of administration of the CSA.

### Student Test Scores

Overall reporting scores for the CSA were produced at the individual student level. To obtain overall reporting scores, the ability (theta) scores needed to be estimated.

ETS Assessment and Learning Technology Research & Development staff reviewed each item and determined the answer keys. The keys were provided to Cambium Assessment, Inc. (CAI) for implementation in the test delivery system (TDS). After CAI finished machine-scoring item responses, scores and responses were delivered to ETS. ETS’ Enterprise Score Key Management (eSKM) system collected and calculated individual students’ overall scores (e.g., total raw scores). ETS’ Psychometric Analysis & Research (PAR) team conducted a series of psychometric analyses such as calibration, equating, and scaling using individual item scores of the test samples and produced the raw-to-scale score conversion tables based on all psychometric analyses. When the conversion tables were produced, eSKM produced the scale score and the score reporting ranges for students who completed the test.

ETS used two parallel scoring systems to produce and verify students’ scores. The eSKM scoring system received individual students’ item scores and item responses from CAI and computed individual student scores for the ETS reporting system. ETS’ PAR team also computed individual student scores based on the same data files using statistical analysis system software. The scores from the two systems were then compared for the purpose of internal quality control. Inconsistency in the total raw scores was investigated and resolved. The parallel scoring process ensured the quality and accuracy of scoring and supported the transfer of scores into the database of the student records scoring system, the Test Operations Management System (TOMS).

#### Theta Scores

Typically, a student’s raw score is the sum of scores on the individual items presented to the student. A theta score is derived from an item response theory (IRT) model and indicates a student’s ability level on the IRT scale. The test for each grade level or the high school grade band—grades three through eight and high school—has its own theta scale.

After all operational items are calibrated and linked onto the base scale, the raw score can be computed as a sum of dichotomous and polytomous item scores and can be transformed into an ability estimate (theta) by using the IRT inverse test characteristic curve (TCC) method (Stocking, 1996). With this method, the student’s estimated ability is the ability value at which the expected raw score is equal to the student’s raw score. Refer to section [*8.4 Item Response Theory Analyses*](#_Item_Response_Theory) for the scaling procedures and the IRT inverse TCC method. Note that the estimation of ability is implemented by using the item parameters of each form.

When a conversion table from the raw score to theta score is created for each form, the theta score of each individual student can be obtained using the conversion table. Refer to [appendix 7.A](#_Appendix_7.A:_Raw) for the score distribution of raw scores, theta scores, and scale scores.

#### Scale Scores

Raw scores obtained on each grade-level CSA are transformed to scale scores using the scaling process described in subsection [*8.4.6 Scaling the Scores*](#_Scaling_the_Scores_1). The following requirements were used to develop and define the CSA reporting scale ranges:

1. Each scale score has three digits (e.g., 320, 551, or 780) where the first digit is indicative of the grade level being reported. The leading digit is defined by the grade level for elementary and middle school, while the high school leading digit is set to “9.” The latter two digits represent the scale score as derived from the transformation from the raw scores to the scale scores. Refer to subsection [*8.4.6.2 Transformation from Theta Scores to Scale Scores*](#_Transformation_from_Theta_1) for details of the transformation.
2. Score ranges are grade-level or grade-band specific. For example, the possible scale scores would be 300 to 399 for grade three with the lowest obtainable scale score (LOSS) at 300 and the highest obtainable scale score (HOSS) at 399. For grade four, this range is 400 to 499 with a LOSS of 400 and a HOSS of 499, and so on for the other grade levels. For high school grade levels, the scale ranges from 900 to 999 with a LOSS of 900 and a HOSS of 999.
3. The scores reported as Range 1, Range 2, and Range 3 on the reporting scale are the same from year to year. Across the grade levels, the last two digits corresponding to the score reporting range are the same, such as 360 for grade three, 460 for grade four, 560 for grade five, 660 for grade six, 760 for grade seven, 860 for grade eight, and 960 for high school. Table 7.4 shows score reporting ranges across grade levels and the high school grade band.
4. Students who logged on to the test and answered at least 1 item, but fewer than 10 items, are assigned the LOSS.

For students who complete a CSA, their scale scores cannot be lower than the LOSS or higher than the HOSS as a result of truncation in the scale score transformation. For example, the scale scores for grade three are truncated at a minimum of 300 and a maximum of 399. As a result, the range of student ability estimates [-6, +6] is transformed to the scale score range [300, 399] for grade three and [400, 499] for grade four. The scale score ranges for other grade levels and the high school grade band follow the same rules.

The complete raw-to-scale score conversion tables for each CSA are presented in table 7.A.1 through table 7.A.7 in [appendix 7.A](#_Appendix_7.A:_Raw). The raw scores, theta scores, transformed scale scores, and the number and percentage of students at each raw score are listed in those tables.

#### Score Reporting Ranges

CSA reporting scales currently classify each student’s performance into one of the three score reporting ranges. Achievement levels will be determined and implemented once the CSA has a more stable test population and different types of programs and curricula have been identified. Detailed information regarding the determination of the score reporting ranges can be found in the *Standard Setting Technical Report for the California Spanish Assessment* (California Department of Education [CDE], 2019). The score reporting ranges for each grade level or high school grade band are presented in table 7.1.

Table 7.1 CSA Score Reporting Ranges by Grade Level or Grade Band

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** |
| Grade 3 | 300–348 | 349–359 | 360–399 |
| Grade 4 | 400–448 | 449–459 | 460–499 |
| Grade 5 | 500–545 | 546–559 | 560–599 |
| Grade 6 | 600–647 | 648–659 | 660–699 |
| Grade 7 | 700–743 | 744–759 | 760–799 |
| Grade 8 | 800–847 | 848–859 | 860–899 |
| High school | 900–949 | 950–959 | 960–999 |

#### Scoring of Incomplete Cases

Whether a test should be scored or reported depended on the “complete” status of the test and how much of the test was submitted for scoring. Depending on the nature of the missing data, different actions were taken.

As defined in the CSA scoring and reporting specifications, for a typical test administration, tests are considered “complete” and students are scored if students respond to at least 10 items. Tests are also considered “complete” with the LOSS; students are assigned the LOSS if students respond to at least 1 but fewer than 10 items. Tests are considered “partially complete,” and no scores are assigned, if students log on to the test but answer no items. Tests are considered “non-complete” if students did not log on to the test.

ETS, in consultation with the CDE, implemented several rules to identify an incomplete test; these rules are represented in table 7.2, which included the following four specifications:

1. Attemptedness and participation rules describing when a test is considered attempted or taken
2. When a test is scored
3. Whether incomplete tests are scored
4. When a score is reported

Table 7.2 Rules for Incomplete Tests

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **If the student** | **Classify the student as taking the test?** | **Score the student’s responses?** | **Classify the student as attempting the test?** | **Report a score for the student?** |
| Logged on to the test and answered at least 1 item but fewer than 10 items | Yes | Yes, LOSS for the test | Yes (Completion with LOSS) | Yes (LOSS) |
| Logged on to the test and answered at least 10 items | Yes | Yes | Yes (Completion) | Yes (Actual scores) |
| Logged on to the test but answered no items | No | N/A | Partial Completion | No |
| Did not log on to the test | No | N/A | Noncompletion | No |
| Logged on and answered at least one item with a special condition code (refer to subsection [*7.3.2 Special Cases*](#_Special_Cases_1)*)* | No | N/A | Not Tested | No |

### Overview of Score Aggregation Procedures

To provide meaningful results to the interest holders, test scores for a given grade level and content area are aggregated at the school, LEA or direct funded charter school, county, and state levels. The aggregated scores are generated both for selected groups and for the population. The next subsection contains a description of the types of aggregation performed on California Assessment of Student Performance and Progress (CAASPP) computer-based assessment scores. Score aggregation includes only students with valid scores; refer to subsection [*7.3.2 Special Cases*](#_Special_Cases_1) for more information.

#### Individual Student Score Distributions and Summary Statistics

Summary statistics that describe student performance are presented in table 7.3 and table 7.4. Included in table 7.3 are the number of students taking each test and the means and standard deviations (SDs) of scale scores and theta scores. In general, the number of students who tested with valid scores in 2021‒22 increased in comparison with the previous years across grade levels.

Table 7.3 Mean and SD of Theta Scores and Scale Scores

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Number of Students Tested with Valid Scores** | **Scale Score Mean** | **Scale Score SD** | **Theta Score Mean** | **Theta Score SD** |
| Grade 3 | 11,105 | 344 | 10.3 | -0.2988 | 0.7376 |
| Grade 4 | 10,119 | 446 | 9.4 | -0.1837 | 0.7318 |
| Grade 5 | 8,434 | 546 | 10.0 | -0.1255 | 0.7134 |
| Grade 6 | 5,892 | 650 | 10.6 | -0.0399 | 0.7601 |
| Grade 7 | 3,833 | 741 | 10.1 | -0.0469 | 0.6779 |
| Grade 8 | 3,239 | 845 | 11.2 | -0.0412 | 0.7519 |
| High school—Grade 9 | 1,736 | 944 | 9.6 | -0.2140 | 0.6892 |
| High school—Grade 10 | 904 | 948 | 9.7 | 0.0553 | 0.6961 |
| High school—Grade 11 | 663 | 947 | 10.6 | -0.0375 | 0.7606 |
| High school—Grade 12 | 210 | 949 | 10.3 | 0.1334 | 0.7371 |
| High school—All grades | 3,513 | 946 | 10.0 | -0.0906 | 0.7192 |

Table 7.4 presents the number and percent of students in the three score reporting ranges. More students are at score reporting range 1 than range 2 or range 3 for all grade levels, and score reporting range 3 has the fewest students. In grades five, six, eight, and ten through twelve, 10 percent or higher of students scored in range 3, and a relatively lower percentage of students scored in range 1, compared to the other grade levels.

Table 7.4 Numbers and Percentages of Students in Score Reporting Ranges

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Range 1 N** | **Range 1 %** | **Range 2 N** | **Range 2 %** | **Range 3 N** | **Range 3 %** |
| Grade 3 | 7,987 | 71.92 | 2,082 | 18.75 | 1,036 | 9.33 |
| Grade 4 | 6,618 | 65.40 | 2,549 | 25.19 | 952 | 9.41 |
| Grade 5 | 4,340 | 51.46 | 3,194 | 37.87 | 900 | 10.67 |
| Grade 6 | 2,726 | 46.27 | 2,076 | 35.23 | 1,090 | 18.50 |
| Grade 7 | 2,368 | 61.78 | 1,251 | 32.64 | 214 | 5.58 |
| Grade 8 | 1,963 | 60.61 | 906 | 27.97 | 370 | 11.42 |
| High school—Grade 9 | 1,268 | 73.04 | 342 | 19.70 | 126 | 7.26 |
| High school—Grade 10 | 510 | 56.42 | 277 | 30.64 | 117 | 12.94 |
| High school—Grade 11 | 425 | 64.10 | 143 | 21.57 | 95 | 14.33 |
| High school—Grade 12 | 105 | 50.00 | 67 | 31.90 | 38 | 18.10 |
| High school—All grades | 2,308 | 65.70 | 829 | 23.60 | 376 | 10.70 |

#### Demographic Student Group Summaries

Statistics summarizing student participation in Spanish instruction programs and their performance by grade level or the high school grade band for selected demographic student groups are provided in [appendix 7.B](#_Appendix_7.B:_Demographic). In table 7.B.1 through table 7.B.7, as well as table 7.C.1 through table 7.C.7 in [appendix 7.C](#_Appendix_7.C:_Demographic_1), students are grouped by demographic characteristics, including gender, ethnicity, English language fluency, economic status, special education services status, length of enrollment in US schools, Spanish-language program type, and percentage of daily instruction in Spanish. For each demographic student group, the number of students who completed testing with a valid reporting scale score, reporting score means and SDs, and the percentage of students in each score reporting range is reported.

Table 7.5 provides definitions of the demographic student groups. To protect student privacy, when the number of students in a student group is 10 or fewer, the summary statistics are not reported and are presented as “N/A.” [Appendix 7.B](#_Appendix_7.B:_Demographic) presents, by demographic student groups, the number of students for Spanish instruction programs and daily instruction in Spanish from the student survey as part of the operational assessment. Table 7.B.1 through table 7.B.7 present the number and percentage of students in Spanish programs by demographic groups. [Appendix 7.C](#_Appendix_7.C:_Demographic_1) comprises the demographic student group performance summaries. Table 7.C.1 through table 7.C.7 present the percentage of students in each score reporting range for total scores by demographic variables for grades three through eight and high school.

Table 7.5 Demographic Student Groups to Be Reported

|  |  |
| --- | --- |
| **Student Group** | **Definition** |
| **Economic Status** | * Not economically disadvantaged * Economically disadvantaged |
| **English Language Fluency** | * English only * Initial fluent English proficient (IFEP) * English learner (EL) * Reclassified fluent English proficient (RFEP) * Adult English learner (ADEL) * Ever-ELs (EL, RFEP, or ADEL) * To be determined * English proficiency unknown |
| **Enrollment in US Schools** | * Less than 12 months * 12 months or more |
| **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 |

Table 7.5 *(continuation)*

|  |  |
| --- | --- |
| **Student Group** | **Definition** |
| **Gender** | * Male * Female * Nonbinary |
| **Percentage of School Day Instruction Provided in Spanish** | * 0–25% * 26–50% * 51–75% * 76–100% |
| **Received Instruction in Spanish in the 2021–22 School Year—Program Type** | * One-Way Immersion * Dual-Language Immersion * Developmental Bilingual * Heritage Language or Indigenous Language * Spanish as a Foreign Language[[4]](#footnote-5) |
| **Special Education Services Status** | * No special education services * Special education services |

### Reports Produced and Scores for Each Report

The tests that make up the CAASPP computer-based assessments provide results or score summaries that are reported for different purposes. The four major purposes are to

1. help facilitate conversations between parents/guardians and teachers about student performance,
2. serve as a tool to help parents/guardians and teachers work together to improve student learning,
3. help schools and LEAs identify strengths and areas that need improvement in their educational programs, and
4. provide the public and policymakers with information about student achievement.

This section provides detailed descriptions of the uses and applications of CAASPP reporting for students.

#### Online Reporting

TOMS is a secure website hosted by ETS that permits LEA users to manage the CAASPP computer-based assessments and to inform the TDS. This system uses a role-specific design to restrict access to certain tools and applications based on the user’s designated role. Specific functions of TOMS include the following:

* Manage user access privileges
* Manage test administration calendars and testing windows
* Manage student test assignments
* Manage and confirm the accuracy of students’ test settings (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 2021–‍22 administration: the California Educator Reporting System (CERS).

TOMS communicated with CERS, which provided authorized users with interactive and cumulative online reports for reading, writing mechanics, and listening at the student, school, group, and LEA levels. CERS provided preliminary score data for each administered test available in the reporting system.

Based on the CAASPP reporting requirements, CERS provided the preliminary summative reports containing information outlining student knowledge and skills. 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 CAASPP results at the LEA, school, grade, classroom, or customized group level. CERS provided these reports, which can be downloaded and used to inform instruction. LEA staff with TOMS logon credentials could enter CERS through the CAASPP 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 was administered out-of-grade level tests.
* The student score was invalidated in the system (not reported in aggregated reporting).

#### Types of Score Reports

There are two categories of CAASPP reports. The specific reports within each category are presented in this subsection.

1. **Student Score Report (SSR)—**The SSR was the official score report for parents/‌guardians. An SSR described the student’s results and was made available only to students who met the program’s participation requirement.
2. **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 CSA SSR is the official score report for parents/guardians and includes the following metrics:

* Reported scale scores (The ranges of scale scores are provided in table 7.1.)
* Reported reporting ranges (CSA reporting ranges are “Score Reporting Range 1,” “Score Reporting Range 2,” and “Score Reporting Range 3.”)
* A description of score reporting ranges

Scores for students who were assigned accommodations or designated supports are reported in the same way as for students who were not assigned accommodations or designated supports. Detailed information about accessibility resources is described in subsection [*5.4.1 Accessibility Resource Categories*](#_Accessibility_Resource_Categories_1).

LEAs had three options for accessing and distributing SSRs to parents/guardians:

1. Accessing electronic SSR PDFs using a locally provided parent/guardian 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 CSA SSR, the LEA sent the printed report(s) to the child’s parent/guardian. CSA SSRs that included individual student results were not distributed beyond the student’s school.

Scores for students who were assigned accommodations or designated supports are reported in the same way as for students who were not assigned accommodations or designated supports. Detailed information about accessibility resources is described in subsection [*5.4.1 Accessibility Resource Categories*](#_Accessibility_Resource_Categories_1).

For the 2021–22 test administration, SSRs were made available to the LEAs in English, Spanish, Filipino, Chinese (Traditional), and Vietnamese. An SSR in a supported language was created if the student’s primary language as reported in the California Longitudinal Pupil Achievement Data System was one of these supported languages. The LEAs that received SSRs in supported languages received one SSR in English and another in the supported language. These reports were available as PDFs for the LEA to download from TOMS.

Further information about the SSR and its interpretation is provided on the CAASPP Starting Smarter website for California assessments.

###### Access via Student or Parent Portal

LEAs had the option to provide SSRs electronically using a locally provided parent or student portal.

Amazon Web Services—with the Amazon Simple Storage Service and the Amazon Key Management Service—ensured encrypted access for parents/guardians to view a child’s electronic SSR, which was available as a PDF.

###### Access via the Test Operations Management System

The LEA CAASPP coordinator downloaded the electronic PDFs directly from TOMS and could 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

CSA results, presented in SSRs, 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 how to identify priorities to help the student progress in Spanish reading/language arts literacy. 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.

There may be a low, moderate, or high degree of alignment between the CSA results and the LEA’s instructional programs. Factors that determine this alignment are as follows:

* Does the LEA’s Spanish language program provide Spanish reading/language arts instruction?
* Is the LEA’s Spanish language program aligned with the California Common Core State Standards en Español?
* Is there a percentage of the LEA’s instructional day that is conducted in Spanish?

If all three statements are true, then an LEA may have a high degree of alignment between its CSA results and its instructional program. The less true the statements are, the lower the degree of the alignment.

With this in mind, schools may use the CSA results to help make decisions about how to support student achievement. CSA results, however, should never be used as the only source of information to make important decisions about a child’s education. CSA results help schools and LEAs identify strengths and weaknesses in their instructional programs.

#### Criteria for Interpreting Test Scores

An LEA may use CAASPP computer-based summative assessment results to help make decisions about student placement, promotion, retention, or other considerations related to student achievement. However, it is important to remember that a single 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 CAASPP computer-based summative assessment results. It is also important to note that a student’s score in a content area could vary somewhat if the student were retested.

#### Criteria for Interpreting Score Reports

The information presented in various reports must be interpreted with caution when making performance comparisons. When comparing scale score and reporting range results, the user is limited to comparisons within a grade level or grade band. The user may compare scale scores for the same grade level, within a school, between schools, or between a school and its LEA, its county, or the state. The CAASPP user can also make comparisons within the same grade level or grade band across years.

However, comparing scale scores from different grade levels for the CAASPP is not appropriate, because the curricula are different across grade levels and the scale scores are not vertically linked between grade levels.

For more details on the criteria for interpreting information provided on the score reports, refer to the CAASPP Starting Smarter website for California assessments or the *2021–22 CAASPP Post-Test Guide* (CDE, 2022), which was applicable for the 2021–22 CAASPP administration.

### References

California Department of Education. (2019). *Standard setting technical report for the California Spanish Assessment.* California Department of Education website.

California Department of Education. (2022). *CAASPP post-test guide: Technical information for student score reports for CAASPP LEA and site coordinators and research specialists.* Sacramento, CA: California Department of Education.

Stocking, M. L. (1996). An alternative method for scoring adaptive tests. *Journal of Educational and Behavioral Statistics, 21,* 365–89.

### Appendix 7.A: Raw Score, Theta Score, Scale Score, and Conditional Standard Error of Measurement Distributions of Students Taking Each Test

**Notes:**

* A student with an incomplete test did not receive a score.
* When a student logged on to the TDS but did not answer any item, the student did not receive a score.
* When a student logged on and answered at least 1 item but fewer than 10 items, the student received the lowest obtainable scale score (such as 300 for grade three, 400 for grade four, …, and 900 for high school grade levels).
* Table 7.A.1 through table 7.A.7 include students who have completed testing with a valid theta score.
* Table 7.A.1 through table 7.A.7 only include students taking the regular forms.

Table 7.A.1 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for Grade Three

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 3 | -3.545 | 300 | 21 | 2 | 0.0 | 2 | 0.0 |
| 4 | -3.212 | 303 | 8 | 1 | 0.0 | 3 | 0.0 |
| 5 | -2.945 | 307 | 7 | 1 | 0.0 | 4 | 0.0 |
| 6 | -2.721 | 310 | 6 | 3 | 0.0 | 7 | 0.1 |
| 7 | -2.527 | 313 | 6 | 2 | 0.0 | 9 | 0.1 |
| 8 | -2.355 | 315 | 6 | 3 | 0.0 | 12 | 0.1 |
| 9 | -2.199 | 317 | 5 | 4 | 0.0 | 16 | 0.2 |
| 10 | -2.057 | 319 | 5 | 1 | 0.0 | 17 | 0.2 |
| 11 | -1.926 | 321 | 5 | 4 | 0.0 | 21 | 0.2 |
| 12 | -1.804 | 323 | 5 | 8 | 0.1 | 29 | 0.3 |
| 13 | -1.690 | 324 | 5 | 19 | 0.2 | 48 | 0.5 |
| 14 | -1.582 | 326 | 5 | 44 | 0.4 | 92 | 0.9 |
| 15 | -1.479 | 327 | 4 | 66 | 0.7 | 158 | 1.6 |
| 16 | -1.382 | 329 | 4 | 108 | 1.1 | 266 | 2.7 |
| 17 | -1.288 | 330 | 4 | 159 | 1.6 | 425 | 4.3 |
| 18 | -1.198 | 331 | 4 | 253 | 2.6 | 678 | 6.9 |
| 19 | -1.111 | 332 | 4 | 322 | 3.3 | 1,000 | 10.2 |
| 20 | -1.027 | 334 | 4 | 416 | 4.2 | 1,416 | 14.4 |
| 21 | -0.946 | 335 | 4 | 530 | 5.4 | 1,946 | 19.8 |
| 22 | -0.866 | 336 | 4 | 495 | 5.0 | 2,441 | 24.8 |

Table 7.A.1 *(continuation one)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 23 | -0.789 | 337 | 4 | 546 | 5.6 | 2,987 | 30.4 |
| 24 | -0.713 | 338 | 4 | 580 | 5.9 | 3,567 | 36.3 |
| 25 | -0.639 | 339 | 4 | 485 | 4.9 | 4,052 | 41.2 |
| 26 | -0.566 | 340 | 4 | 479 | 4.9 | 4,531 | 46.1 |
| 27 | -0.494 | 341 | 4 | 434 | 4.4 | 4,965 | 50.5 |
| 28 | -0.422 | 342 | 4 | 365 | 3.7 | 5,330 | 54.2 |
| 29 | -0.352 | 343 | 4 | 334 | 3.4 | 5,664 | 57.6 |
| 30 | -0.282 | 344 | 4 | 325 | 3.3 | 5,989 | 60.9 |
| 31 | -0.213 | 345 | 4 | 312 | 3.2 | 6,301 | 64.1 |
| 32 | -0.144 | 346 | 4 | 283 | 2.9 | 6,584 | 67.0 |
| 33 | -0.076 | 347 | 4 | 199 | 2.0 | 6,783 | 69.0 |
| 34 | -0.007 | 348 | 4 | 230 | 2.3 | 7,013 | 71.4 |
| 35 | 0.061 | 349 | 4 | 211 | 2.1 | 7,224 | 73.5 |
| 36 | 0.129 | 350 | 4 | 203 | 2.1 | 7,427 | 75.6 |
| 37 | 0.198 | 351 | 4 | 173 | 1.8 | 7,600 | 77.3 |
| 38 | 0.267 | 352 | 4 | 166 | 1.7 | 7,766 | 79.0 |
| 39 | 0.337 | 353 | 4 | 190 | 1.9 | 7,956 | 81.0 |
| 40 | 0.407 | 354 | 4 | 179 | 1.8 | 8,135 | 82.8 |
| 41 | 0.477 | 355 | 4 | 159 | 1.6 | 8,294 | 84.4 |
| 42 | 0.549 | 356 | 4 | 145 | 1.5 | 8,439 | 85.9 |
| 43 | 0.622 | 357 | 4 | 145 | 1.5 | 8,584 | 87.3 |
| 44 | 0.696 | 358 | 4 | 143 | 1.5 | 8,727 | 88.8 |
| 45 | 0.771 | 359 | 4 | 127 | 1.3 | 8,854 | 90.1 |
| 46 | 0.848 | 360 | 4 | 130 | 1.3 | 8,984 | 91.4 |
| 47 | 0.927 | 361 | 4 | 105 | 1.1 | 9,089 | 92.5 |
| 48 | 1.008 | 362 | 4 | 100 | 1.0 | 9,189 | 93.5 |
| 49 | 1.092 | 363 | 4 | 98 | 1.0 | 9,287 | 94.5 |
| 50 | 1.179 | 365 | 4 | 92 | 0.9 | 9,379 | 95.4 |
| 51 | 1.269 | 366 | 4 | 78 | 0.8 | 9,457 | 96.2 |
| 52 | 1.363 | 367 | 4 | 74 | 0.8 | 9,531 | 97.0 |
| 53 | 1.462 | 368 | 4 | 64 | 0.7 | 9,595 | 97.6 |
| 54 | 1.566 | 370 | 5 | 57 | 0.6 | 9,652 | 98.2 |
| 55 | 1.677 | 371 | 5 | 48 | 0.5 | 9,700 | 98.7 |
| 56 | 1.796 | 373 | 5 | 36 | 0.4 | 9,736 | 99.1 |
| 57 | 1.924 | 375 | 5 | 32 | 0.3 | 9,768 | 99.4 |
| 58 | 2.065 | 377 | 5 | 21 | 0.2 | 9,789 | 99.6 |
| 59 | 2.221 | 379 | 6 | 19 | 0.2 | 9,808 | 99.8 |

Table 7.A.1 *(continuation two)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 60 | 2.398 | 382 | 6 | 9 | 0.1 | 9,817 | 99.9 |
| 61 | 2.602 | 384 | 7 | 7 | 0.1 | 9,824 | 100.0 |
| 62 | 2.847 | 388 | 7 | 1 | 0.0 | 9,825 | 100.0 |
| 63 | 3.156 | 392 | 8 | 2 | 0.0 | 9,827 | 100.0 |
| 64 | 3.583 | 398 | 10 | 1 | 0.0 | 9,828 | 100.0 |

Table 7.A.2 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for Grade Four

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 2 | -4.075 | 400 | 20 | 1 | 0.0 | 1 | 0.0 |
| 3 | -3.623 | 401 | 8 | 1 | 0.0 | 2 | 0.0 |
| 8 | -2.424 | 416 | 5 | 1 | 0.0 | 3 | 0.0 |
| 11 | -1.984 | 422 | 5 | 2 | 0.0 | 5 | 0.1 |
| 12 | -1.858 | 424 | 5 | 8 | 0.1 | 13 | 0.1 |
| 13 | -1.739 | 425 | 4 | 8 | 0.1 | 21 | 0.2 |
| 14 | -1.626 | 427 | 4 | 23 | 0.3 | 44 | 0.5 |
| 15 | -1.518 | 428 | 4 | 37 | 0.4 | 81 | 0.9 |
| 16 | -1.416 | 430 | 4 | 73 | 0.8 | 154 | 1.7 |
| 17 | -1.317 | 431 | 4 | 82 | 0.9 | 236 | 2.6 |
| 18 | -1.222 | 432 | 4 | 142 | 1.5 | 378 | 4.1 |
| 19 | -1.130 | 433 | 4 | 211 | 2.3 | 589 | 6.4 |
| 20 | -1.041 | 434 | 4 | 283 | 3.1 | 872 | 9.5 |
| 21 | -0.954 | 436 | 4 | 367 | 4.0 | 1,239 | 13.5 |
| 22 | -0.869 | 437 | 4 | 398 | 4.3 | 1,637 | 17.9 |
| 23 | -0.787 | 438 | 4 | 437 | 4.8 | 2,074 | 22.6 |
| 24 | -0.706 | 439 | 4 | 455 | 5.0 | 2,529 | 27.6 |
| 25 | -0.626 | 440 | 4 | 454 | 5.0 | 2,983 | 32.6 |
| 26 | -0.548 | 441 | 4 | 496 | 5.4 | 3,479 | 38.0 |
| 27 | -0.472 | 442 | 4 | 400 | 4.4 | 3,879 | 42.3 |
| 28 | -0.396 | 443 | 4 | 413 | 4.5 | 4,292 | 46.8 |
| 29 | -0.321 | 444 | 4 | 407 | 4.4 | 4,699 | 51.3 |
| 30 | -0.247 | 445 | 4 | 368 | 4.0 | 5,067 | 55.3 |
| 31 | -0.173 | 446 | 4 | 330 | 3.6 | 5,397 | 58.9 |
| 32 | -0.100 | 447 | 4 | 287 | 3.1 | 5,684 | 62.0 |
| 33 | -0.027 | 448 | 4 | 276 | 3.0 | 5,960 | 65.0 |
| 34 | 0.045 | 449 | 3 | 257 | 2.8 | 6,217 | 67.8 |
| 35 | 0.118 | 450 | 3 | 271 | 3.0 | 6,488 | 70.8 |
| 36 | 0.190 | 450 | 3 | 235 | 2.6 | 6,723 | 73.4 |
| 37 | 0.263 | 451 | 4 | 235 | 2.6 | 6,958 | 75.9 |
| 38 | 0.336 | 452 | 4 | 221 | 2.4 | 7,179 | 78.3 |
| 39 | 0.409 | 453 | 4 | 186 | 2.0 | 7,365 | 80.4 |
| 40 | 0.483 | 454 | 4 | 172 | 1.9 | 7,537 | 82.3 |
| 41 | 0.558 | 455 | 4 | 182 | 2.0 | 7,719 | 84.2 |
| 42 | 0.633 | 456 | 4 | 148 | 1.6 | 7,867 | 85.9 |
| 43 | 0.710 | 457 | 4 | 161 | 1.8 | 8,028 | 87.6 |
| 44 | 0.788 | 458 | 4 | 149 | 1.6 | 8,177 | 89.2 |

Table 7.A.2 *(continuation)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 45 | 0.867 | 459 | 4 | 112 | 1.2 | 8,289 | 90.5 |
| 46 | 0.948 | 460 | 4 | 130 | 1.4 | 8,419 | 91.9 |
| 47 | 1.030 | 461 | 4 | 112 | 1.2 | 8,531 | 93.1 |
| 48 | 1.116 | 463 | 4 | 106 | 1.2 | 8,637 | 94.3 |
| 49 | 1.203 | 464 | 4 | 99 | 1.1 | 8,736 | 95.3 |
| 50 | 1.294 | 465 | 4 | 83 | 0.9 | 8,819 | 96.2 |
| 51 | 1.388 | 466 | 4 | 71 | 0.8 | 8,890 | 97.0 |
| 52 | 1.486 | 467 | 4 | 65 | 0.7 | 8,955 | 97.7 |
| 53 | 1.589 | 469 | 4 | 52 | 0.6 | 9,007 | 98.3 |
| 54 | 1.698 | 470 | 4 | 47 | 0.5 | 9,054 | 98.8 |
| 55 | 1.813 | 472 | 4 | 32 | 0.3 | 9,086 | 99.2 |
| 56 | 1.936 | 473 | 5 | 37 | 0.4 | 9,123 | 99.6 |
| 57 | 2.068 | 475 | 5 | 15 | 0.2 | 9,138 | 99.7 |
| 58 | 2.213 | 477 | 5 | 10 | 0.1 | 9,148 | 99.8 |
| 59 | 2.374 | 479 | 5 | 9 | 0.1 | 9,157 | 99.9 |
| 60 | 2.554 | 481 | 6 | 3 | 0.0 | 9,160 | 100.0 |
| 61 | 2.763 | 484 | 6 | 2 | 0.0 | 9,162 | 100.0 |
| 63 | 3.326 | 491 | 8 | 1 | 0.0 | 9,163 | 100.0 |

Table 7.A.3 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for Grade Five

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 8 | -2.253 | 516 | 6 | 1 | 0.0 | 1 | 0.0 |
| 9 | -2.095 | 519 | 5 | 2 | 0.0 | 3 | 0.0 |
| 10 | -1.950 | 521 | 5 | 3 | 0.0 | 6 | 0.1 |
| 11 | -1.816 | 523 | 5 | 7 | 0.1 | 13 | 0.2 |
| 12 | -1.691 | 524 | 5 | 17 | 0.2 | 30 | 0.4 |
| 13 | -1.573 | 526 | 5 | 26 | 0.3 | 56 | 0.7 |
| 14 | -1.461 | 528 | 5 | 37 | 0.5 | 93 | 1.2 |
| 15 | -1.354 | 529 | 5 | 83 | 1.1 | 176 | 2.3 |
| 16 | -1.253 | 530 | 4 | 94 | 1.2 | 270 | 3.5 |
| 17 | -1.155 | 532 | 4 | 150 | 1.9 | 420 | 5.4 |
| 18 | -1.060 | 533 | 4 | 207 | 2.7 | 627 | 8.1 |
| 19 | -0.969 | 534 | 4 | 245 | 3.2 | 872 | 11.2 |
| 20 | -0.880 | 536 | 4 | 292 | 3.8 | 1,164 | 15.0 |
| 21 | -0.794 | 537 | 4 | 366 | 4.7 | 1,530 | 19.7 |
| 22 | -0.710 | 538 | 4 | 361 | 4.7 | 1,891 | 24.4 |
| 23 | -0.627 | 539 | 4 | 356 | 4.6 | 2,247 | 28.9 |
| 24 | -0.547 | 540 | 4 | 380 | 4.9 | 2,627 | 33.8 |
| 25 | -0.467 | 541 | 4 | 360 | 4.6 | 2,987 | 38.5 |
| 26 | -0.389 | 543 | 4 | 335 | 4.3 | 3,322 | 42.8 |
| 27 | -0.312 | 544 | 4 | 297 | 3.8 | 3,619 | 46.6 |
| 28 | -0.236 | 545 | 4 | 319 | 4.1 | 3,938 | 50.7 |
| 29 | -0.161 | 546 | 4 | 317 | 4.1 | 4,255 | 54.8 |
| 30 | -0.086 | 547 | 4 | 265 | 3.4 | 4,520 | 58.2 |
| 31 | -0.012 | 548 | 4 | 276 | 3.6 | 4,796 | 61.8 |
| 32 | 0.062 | 549 | 4 | 244 | 3.1 | 5,040 | 64.9 |
| 33 | 0.136 | 550 | 4 | 266 | 3.4 | 5,306 | 68.3 |
| 34 | 0.210 | 551 | 4 | 219 | 2.8 | 5,525 | 71.2 |
| 35 | 0.284 | 552 | 4 | 248 | 3.2 | 5,773 | 74.4 |
| 36 | 0.358 | 553 | 4 | 189 | 2.4 | 5,962 | 76.8 |
| 37 | 0.432 | 554 | 4 | 183 | 2.4 | 6,145 | 79.2 |
| 38 | 0.507 | 555 | 4 | 176 | 2.3 | 6,321 | 81.4 |
| 39 | 0.582 | 556 | 4 | 148 | 1.9 | 6,469 | 83.3 |
| 40 | 0.659 | 557 | 4 | 138 | 1.8 | 6,607 | 85.1 |
| 41 | 0.736 | 558 | 4 | 153 | 2.0 | 6,760 | 87.1 |
| 42 | 0.814 | 559 | 4 | 147 | 1.9 | 6,907 | 89.0 |
| 43 | 0.894 | 561 | 4 | 134 | 1.7 | 7,041 | 90.7 |
| 44 | 0.976 | 562 | 4 | 112 | 1.4 | 7,153 | 92.1 |

Table 7.A.3 *(continuation)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 45 | 1.059 | 563 | 4 | 106 | 1.4 | 7,259 | 93.5 |
| 46 | 1.144 | 564 | 4 | 104 | 1.3 | 7,363 | 94.8 |
| 47 | 1.232 | 565 | 4 | 78 | 1.0 | 7,441 | 95.9 |
| 48 | 1.323 | 567 | 4 | 83 | 1.1 | 7,524 | 96.9 |
| 49 | 1.418 | 568 | 4 | 57 | 0.7 | 7,581 | 97.7 |
| 50 | 1.516 | 569 | 4 | 43 | 0.6 | 7,624 | 98.2 |
| 51 | 1.619 | 571 | 5 | 40 | 0.5 | 7,664 | 98.7 |
| 52 | 1.728 | 572 | 5 | 32 | 0.4 | 7,696 | 99.1 |
| 53 | 1.843 | 574 | 5 | 29 | 0.4 | 7,725 | 99.5 |
| 54 | 1.965 | 576 | 5 | 18 | 0.2 | 7,743 | 99.7 |
| 55 | 2.098 | 577 | 5 | 9 | 0.1 | 7,752 | 99.9 |
| 56 | 2.242 | 579 | 5 | 3 | 0.0 | 7,755 | 99.9 |
| 57 | 2.402 | 582 | 6 | 4 | 0.1 | 7,759 | 99.9 |
| 58 | 2.583 | 584 | 6 | 2 | 0.0 | 7,761 | 100.0 |
| 59 | 2.791 | 587 | 7 | 1 | 0.0 | 7,762 | 100.0 |
| 60 | 3.039 | 591 | 7 | 1 | 0.0 | 7,763 | 100.0 |

Table 7.A.4 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for Grade Six

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 3 | -3.542 | 600 | 21 | 1 | 0.0 | 1 | 0.0 |
| 4 | -3.215 | 605 | 8 | 1 | 0.0 | 2 | 0.0 |
| 7 | -2.546 | 614 | 6 | 1 | 0.0 | 3 | 0.1 |
| 8 | -2.377 | 617 | 6 | 1 | 0.0 | 4 | 0.1 |
| 9 | -2.225 | 619 | 5 | 3 | 0.1 | 7 | 0.1 |
| 10 | -2.086 | 621 | 5 | 1 | 0.0 | 8 | 0.1 |
| 11 | -1.958 | 623 | 5 | 6 | 0.1 | 14 | 0.3 |
| 12 | -1.838 | 624 | 5 | 4 | 0.1 | 18 | 0.3 |
| 13 | -1.725 | 626 | 5 | 10 | 0.2 | 28 | 0.5 |
| 14 | -1.618 | 627 | 5 | 10 | 0.2 | 38 | 0.7 |
| 15 | -1.516 | 629 | 4 | 19 | 0.4 | 57 | 1.1 |
| 16 | -1.418 | 630 | 4 | 38 | 0.7 | 95 | 1.8 |
| 17 | -1.324 | 631 | 4 | 46 | 0.8 | 141 | 2.6 |
| 18 | -1.234 | 633 | 4 | 78 | 1.4 | 219 | 4.0 |
| 19 | -1.147 | 634 | 4 | 95 | 1.8 | 314 | 5.8 |
| 20 | -1.062 | 635 | 4 | 134 | 2.5 | 448 | 8.3 |
| 21 | -0.979 | 636 | 4 | 139 | 2.6 | 587 | 10.8 |
| 22 | -0.898 | 637 | 4 | 178 | 3.3 | 765 | 14.1 |
| 23 | -0.819 | 639 | 4 | 178 | 3.3 | 943 | 17.4 |
| 24 | -0.742 | 640 | 4 | 210 | 3.9 | 1,153 | 21.3 |
| 25 | -0.666 | 641 | 4 | 167 | 3.1 | 1,320 | 24.4 |
| 26 | -0.591 | 642 | 4 | 201 | 3.7 | 1,521 | 28.1 |
| 27 | -0.516 | 643 | 4 | 190 | 3.5 | 1,711 | 31.6 |
| 28 | -0.443 | 644 | 4 | 204 | 3.8 | 1,915 | 35.4 |
| 29 | -0.371 | 645 | 4 | 194 | 3.6 | 2,109 | 38.9 |
| 30 | -0.298 | 646 | 4 | 199 | 3.7 | 2,308 | 42.6 |
| 31 | -0.227 | 647 | 4 | 195 | 3.6 | 2,503 | 46.2 |
| 32 | -0.156 | 648 | 4 | 167 | 3.1 | 2,670 | 49.3 |
| 33 | -0.084 | 649 | 4 | 156 | 2.9 | 2,826 | 52.2 |
| 34 | -0.013 | 650 | 4 | 171 | 3.2 | 2,997 | 55.3 |
| 35 | 0.058 | 651 | 4 | 180 | 3.3 | 3,177 | 58.7 |
| 36 | 0.129 | 652 | 4 | 178 | 3.3 | 3,355 | 61.9 |
| 37 | 0.201 | 653 | 4 | 168 | 3.1 | 3,523 | 65.0 |
| 38 | 0.273 | 654 | 4 | 174 | 3.2 | 3,697 | 68.3 |
| 39 | 0.346 | 655 | 4 | 157 | 2.9 | 3,854 | 71.2 |
| 40 | 0.420 | 656 | 4 | 139 | 2.6 | 3,993 | 73.7 |
| 41 | 0.494 | 657 | 4 | 141 | 2.6 | 4,134 | 76.3 |

Table 7.A.4 *(continuation)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 42 | 0.570 | 658 | 4 | 151 | 2.8 | 4,285 | 79.1 |
| 43 | 0.646 | 659 | 4 | 117 | 2.2 | 4,402 | 81.3 |
| 44 | 0.725 | 660 | 4 | 122 | 2.3 | 4,524 | 83.5 |
| 45 | 0.804 | 661 | 4 | 130 | 2.4 | 4,654 | 85.9 |
| 46 | 0.886 | 662 | 4 | 124 | 2.3 | 4,778 | 88.2 |
| 47 | 0.971 | 664 | 4 | 112 | 2.1 | 4,890 | 90.3 |
| 48 | 1.057 | 665 | 4 | 92 | 1.7 | 4,982 | 92.0 |
| 49 | 1.147 | 666 | 4 | 87 | 1.6 | 5,069 | 93.6 |
| 50 | 1.240 | 667 | 4 | 74 | 1.4 | 5,143 | 95.0 |
| 51 | 1.337 | 669 | 4 | 73 | 1.3 | 5,216 | 96.3 |
| 52 | 1.439 | 670 | 5 | 60 | 1.1 | 5,276 | 97.4 |
| 53 | 1.547 | 672 | 5 | 29 | 0.5 | 5,305 | 98.0 |
| 54 | 1.661 | 673 | 5 | 33 | 0.6 | 5,338 | 98.6 |
| 55 | 1.783 | 675 | 5 | 26 | 0.5 | 5,364 | 99.0 |
| 56 | 1.915 | 677 | 5 | 20 | 0.4 | 5,384 | 99.4 |
| 57 | 2.058 | 679 | 5 | 9 | 0.2 | 5,393 | 99.6 |
| 58 | 2.218 | 681 | 6 | 13 | 0.2 | 5,406 | 99.8 |
| 59 | 2.397 | 684 | 6 | 4 | 0.1 | 5,410 | 99.9 |
| 60 | 2.604 | 686 | 7 | 3 | 0.1 | 5,413 | 99.9 |
| 61 | 2.852 | 690 | 7 | 1 | 0.0 | 5,414 | 100.0 |
| 62 | 3.164 | 694 | 8 | 2 | 0.0 | 5,416 | 100.0 |

Table 7.A.5 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for Grade Seven

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 3 | -3.389 | 700 | 23 | 2 | 0.1 | 2 | 0.1 |
| 4 | -3.059 | 700 | 23 | 2 | 0.1 | 4 | 0.1 |
| 5 | -2.794 | 700 | 23 | 1 | 0.0 | 5 | 0.1 |
| 7 | -2.379 | 706 | 6 | 1 | 0.0 | 6 | 0.2 |
| 8 | -2.207 | 709 | 6 | 4 | 0.1 | 10 | 0.3 |
| 10 | -1.909 | 713 | 6 | 3 | 0.1 | 13 | 0.3 |
| 11 | -1.777 | 715 | 5 | 2 | 0.1 | 15 | 0.4 |
| 12 | -1.654 | 717 | 5 | 1 | 0.0 | 16 | 0.4 |
| 13 | -1.539 | 719 | 5 | 8 | 0.2 | 24 | 0.6 |
| 14 | -1.429 | 721 | 5 | 6 | 0.2 | 30 | 0.8 |
| 15 | -1.326 | 722 | 5 | 14 | 0.4 | 44 | 1.2 |
| 16 | -1.226 | 724 | 5 | 28 | 0.8 | 72 | 1.9 |
| 17 | -1.131 | 725 | 5 | 35 | 0.9 | 107 | 2.9 |
| 18 | -1.040 | 726 | 4 | 69 | 1.9 | 176 | 4.7 |
| 19 | -0.951 | 728 | 4 | 99 | 2.7 | 275 | 7.4 |
| 20 | -0.865 | 729 | 4 | 108 | 2.9 | 383 | 10.3 |
| 21 | -0.782 | 730 | 4 | 156 | 4.2 | 539 | 14.5 |
| 22 | -0.700 | 732 | 4 | 140 | 3.8 | 679 | 18.2 |
| 23 | -0.621 | 733 | 4 | 140 | 3.8 | 819 | 22.0 |
| 24 | -0.543 | 734 | 4 | 178 | 4.8 | 997 | 26.8 |
| 25 | -0.467 | 735 | 4 | 173 | 4.6 | 1,170 | 31.4 |
| 26 | -0.391 | 736 | 4 | 177 | 4.8 | 1,347 | 36.2 |
| 27 | -0.317 | 737 | 4 | 168 | 4.5 | 1,515 | 40.7 |
| 28 | -0.244 | 738 | 4 | 182 | 4.9 | 1,697 | 45.6 |
| 29 | -0.171 | 739 | 4 | 177 | 4.8 | 1,874 | 50.3 |
| 30 | -0.099 | 741 | 4 | 154 | 4.1 | 2,028 | 54.4 |
| 31 | -0.027 | 742 | 4 | 135 | 3.6 | 2,163 | 58.1 |
| 32 | 0.044 | 743 | 4 | 140 | 3.8 | 2,303 | 61.8 |
| 33 | 0.115 | 744 | 4 | 128 | 3.4 | 2,431 | 65.3 |
| 34 | 0.186 | 745 | 4 | 105 | 2.8 | 2,536 | 68.1 |
| 35 | 0.257 | 746 | 4 | 121 | 3.2 | 2,657 | 71.3 |
| 36 | 0.329 | 747 | 4 | 107 | 2.9 | 2,764 | 74.2 |
| 37 | 0.401 | 748 | 4 | 103 | 2.8 | 2,867 | 77.0 |
| 38 | 0.473 | 749 | 4 | 96 | 2.6 | 2,963 | 79.5 |
| 39 | 0.546 | 750 | 4 | 86 | 2.3 | 3,049 | 81.9 |
| 40 | 0.620 | 751 | 4 | 86 | 2.3 | 3,135 | 84.2 |
| 41 | 0.695 | 752 | 4 | 68 | 1.8 | 3,203 | 86.0 |

Table 7.A.5 *(continuation)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 42 | 0.771 | 754 | 4 | 78 | 2.1 | 3,281 | 88.1 |
| 43 | 0.848 | 755 | 4 | 76 | 2.0 | 3,357 | 90.1 |
| 44 | 0.927 | 756 | 4 | 59 | 1.6 | 3,416 | 91.7 |
| 45 | 1.008 | 757 | 4 | 55 | 1.5 | 3,471 | 93.2 |
| 46 | 1.091 | 758 | 4 | 45 | 1.2 | 3,516 | 94.4 |
| 47 | 1.176 | 760 | 4 | 43 | 1.2 | 3,559 | 95.5 |
| 48 | 1.264 | 761 | 4 | 33 | 0.9 | 3,592 | 96.4 |
| 49 | 1.355 | 762 | 5 | 26 | 0.7 | 3,618 | 97.1 |
| 50 | 1.450 | 764 | 5 | 30 | 0.8 | 3,648 | 97.9 |
| 51 | 1.550 | 765 | 5 | 29 | 0.8 | 3,677 | 98.7 |
| 52 | 1.654 | 767 | 5 | 14 | 0.4 | 3,691 | 99.1 |
| 53 | 1.765 | 768 | 5 | 10 | 0.3 | 3,701 | 99.4 |
| 54 | 1.882 | 770 | 5 | 11 | 0.3 | 3,712 | 99.7 |
| 55 | 2.009 | 772 | 5 | 3 | 0.1 | 3,715 | 99.7 |
| 56 | 2.145 | 774 | 6 | 6 | 0.2 | 3,721 | 99.9 |
| 57 | 2.295 | 776 | 6 | 3 | 0.1 | 3,724 | 100.0 |
| 58 | 2.462 | 779 | 6 | 1 | 0.0 | 3,725 | 100.0 |

Table 7.A.6 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for Grade Eight

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 1 | -4.757 | 800 | 23 | 1 | 0.0 | 1 | 0.0 |
| 5 | -2.953 | 802 | 7 | 1 | 0.0 | 2 | 0.1 |
| 6 | -2.726 | 805 | 7 | 1 | 0.0 | 3 | 0.1 |
| 8 | -2.353 | 811 | 6 | 1 | 0.0 | 4 | 0.1 |
| 10 | -2.048 | 815 | 6 | 3 | 0.1 | 7 | 0.2 |
| 11 | -1.913 | 817 | 5 | 4 | 0.1 | 11 | 0.3 |
| 12 | -1.786 | 819 | 5 | 3 | 0.1 | 14 | 0.4 |
| 13 | -1.668 | 821 | 5 | 5 | 0.2 | 19 | 0.6 |
| 14 | -1.555 | 823 | 5 | 5 | 0.2 | 24 | 0.8 |
| 15 | -1.448 | 824 | 5 | 15 | 0.5 | 39 | 1.2 |
| 16 | -1.345 | 826 | 5 | 20 | 0.6 | 59 | 1.9 |
| 17 | -1.247 | 827 | 5 | 32 | 1.0 | 91 | 2.9 |
| 18 | -1.152 | 829 | 5 | 55 | 1.7 | 146 | 4.6 |
| 19 | -1.061 | 830 | 5 | 60 | 1.9 | 206 | 6.5 |
| 20 | -0.972 | 831 | 4 | 75 | 2.4 | 281 | 8.9 |
| 21 | -0.886 | 833 | 4 | 113 | 3.6 | 394 | 12.5 |
| 22 | -0.801 | 834 | 4 | 101 | 3.2 | 495 | 15.7 |
| 23 | -0.719 | 835 | 4 | 147 | 4.7 | 642 | 20.4 |
| 24 | -0.638 | 836 | 4 | 153 | 4.9 | 795 | 25.2 |
| 25 | -0.559 | 838 | 4 | 116 | 3.7 | 911 | 28.9 |
| 26 | -0.481 | 839 | 4 | 145 | 4.6 | 1,056 | 33.5 |
| 27 | -0.404 | 840 | 4 | 149 | 4.7 | 1,205 | 38.2 |
| 28 | -0.329 | 841 | 4 | 129 | 4.1 | 1,334 | 42.3 |
| 29 | -0.253 | 842 | 4 | 130 | 4.1 | 1,464 | 46.4 |
| 30 | -0.179 | 843 | 4 | 111 | 3.5 | 1,575 | 49.9 |
| 31 | -0.105 | 844 | 4 | 122 | 3.9 | 1,697 | 53.8 |
| 32 | -0.031 | 846 | 4 | 114 | 3.6 | 1,811 | 57.4 |
| 33 | 0.042 | 847 | 4 | 99 | 3.1 | 1,910 | 60.6 |
| 34 | 0.115 | 848 | 4 | 98 | 3.1 | 2,008 | 63.7 |
| 35 | 0.189 | 849 | 4 | 94 | 3.0 | 2,102 | 66.6 |
| 36 | 0.262 | 850 | 4 | 91 | 2.9 | 2,193 | 69.5 |
| 37 | 0.336 | 851 | 4 | 93 | 2.9 | 2,286 | 72.5 |
| 38 | 0.411 | 852 | 4 | 81 | 2.6 | 2,367 | 75.0 |
| 39 | 0.486 | 853 | 4 | 68 | 2.2 | 2,435 | 77.2 |
| 40 | 0.562 | 854 | 4 | 79 | 2.5 | 2,514 | 79.7 |
| 41 | 0.639 | 856 | 4 | 74 | 2.3 | 2,588 | 82.1 |
| 42 | 0.717 | 857 | 4 | 62 | 2.0 | 2,650 | 84.0 |

Table 7.A.6 *(continuation)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 43 | 0.796 | 858 | 4 | 65 | 2.1 | 2,715 | 86.1 |
| 44 | 0.877 | 859 | 4 | 75 | 2.4 | 2,790 | 88.5 |
| 45 | 0.960 | 860 | 4 | 61 | 1.9 | 2,851 | 90.4 |
| 46 | 1.046 | 862 | 4 | 35 | 1.1 | 2,886 | 91.5 |
| 47 | 1.133 | 863 | 4 | 40 | 1.3 | 2,926 | 92.8 |
| 48 | 1.224 | 864 | 5 | 47 | 1.5 | 2,973 | 94.3 |
| 49 | 1.318 | 866 | 5 | 38 | 1.2 | 3,011 | 95.5 |
| 50 | 1.416 | 867 | 5 | 36 | 1.1 | 3,047 | 96.6 |
| 51 | 1.519 | 869 | 5 | 25 | 0.8 | 3,072 | 97.4 |
| 52 | 1.628 | 870 | 5 | 25 | 0.8 | 3,097 | 98.2 |
| 53 | 1.743 | 872 | 5 | 22 | 0.7 | 3,119 | 98.9 |
| 54 | 1.866 | 874 | 5 | 8 | 0.3 | 3,127 | 99.1 |
| 55 | 1.998 | 876 | 6 | 11 | 0.3 | 3,138 | 99.5 |
| 56 | 2.143 | 878 | 6 | 5 | 0.2 | 3,143 | 99.7 |
| 57 | 2.303 | 881 | 6 | 2 | 0.1 | 3,145 | 99.7 |
| 58 | 2.484 | 883 | 7 | 7 | 0.2 | 3,152 | 99.9 |
| 59 | 2.692 | 886 | 7 | 2 | 0.1 | 3,154 | 100.0 |

Table 7.A.7 Overall Raw Score, Theta Score, Scale Score, and CSEM Distribution for High School

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 3 | -3.252 | 901 | 9 | 2 | 0.1 | 2 | 0.1 |
| 4 | -2.916 | 906 | 8 | 2 | 0.1 | 4 | 0.1 |
| 5 | -2.648 | 910 | 7 | 2 | 0.1 | 6 | 0.2 |
| 6 | -2.424 | 913 | 6 | 3 | 0.1 | 9 | 0.3 |
| 7 | -2.229 | 916 | 6 | 2 | 0.1 | 11 | 0.3 |
| 8 | -2.057 | 918 | 6 | 4 | 0.1 | 15 | 0.4 |
| 9 | -1.901 | 920 | 5 | 2 | 0.1 | 17 | 0.5 |
| 10 | -1.759 | 922 | 5 | 3 | 0.1 | 20 | 0.6 |
| 11 | -1.627 | 924 | 5 | 13 | 0.4 | 33 | 0.9 |
| 12 | -1.505 | 926 | 5 | 15 | 0.4 | 48 | 1.4 |
| 13 | -1.389 | 928 | 5 | 17 | 0.5 | 65 | 1.9 |
| 14 | -1.280 | 929 | 5 | 34 | 1.0 | 99 | 2.8 |
| 15 | -1.177 | 931 | 4 | 70 | 2.0 | 169 | 4.8 |
| 16 | -1.078 | 932 | 4 | 91 | 2.6 | 260 | 7.4 |
| 17 | -0.982 | 933 | 4 | 110 | 3.1 | 370 | 10.6 |
| 18 | -0.891 | 935 | 4 | 119 | 3.4 | 489 | 14.0 |
| 19 | -0.802 | 936 | 4 | 133 | 3.8 | 622 | 17.8 |
| 20 | -0.716 | 937 | 4 | 150 | 4.3 | 772 | 22.1 |
| 21 | -0.632 | 938 | 4 | 148 | 4.2 | 920 | 26.3 |
| 22 | -0.549 | 939 | 4 | 159 | 4.5 | 1,079 | 30.8 |
| 23 | -0.469 | 940 | 4 | 162 | 4.6 | 1,241 | 35.5 |
| 24 | -0.390 | 942 | 4 | 154 | 4.4 | 1,395 | 39.9 |
| 25 | -0.312 | 943 | 4 | 144 | 4.1 | 1,539 | 44.0 |
| 26 | -0.236 | 944 | 4 | 156 | 4.5 | 1,695 | 48.4 |
| 27 | -0.160 | 945 | 4 | 135 | 3.9 | 1,830 | 52.3 |
| 28 | -0.085 | 946 | 4 | 125 | 3.6 | 1,955 | 55.9 |
| 29 | -0.010 | 947 | 4 | 124 | 3.5 | 2,079 | 59.4 |
| 30 | 0.064 | 948 | 4 | 99 | 2.8 | 2,178 | 62.2 |
| 31 | 0.138 | 949 | 4 | 124 | 3.5 | 2,302 | 65.8 |
| 32 | 0.212 | 950 | 4 | 99 | 2.8 | 2,401 | 68.6 |
| 33 | 0.286 | 951 | 4 | 106 | 3.0 | 2,507 | 71.6 |
| 34 | 0.360 | 952 | 4 | 85 | 2.4 | 2,592 | 74.1 |
| 35 | 0.434 | 953 | 4 | 117 | 3.3 | 2,709 | 77.4 |
| 36 | 0.509 | 954 | 4 | 100 | 2.9 | 2,809 | 80.3 |
| 37 | 0.585 | 955 | 4 | 99 | 2.8 | 2,908 | 83.1 |
| 38 | 0.661 | 956 | 4 | 75 | 2.1 | 2,983 | 85.3 |
| 39 | 0.739 | 957 | 4 | 85 | 2.4 | 3,068 | 87.7 |

Table 7.A.7 *(continuation)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Raw Score** | **Theta Score** | **Scale Score** | **CSEM** | **N** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 40 | 0.817 | 958 | 4 | 59 | 1.7 | 3,127 | 89.4 |
| 41 | 0.897 | 960 | 4 | 64 | 1.8 | 3,191 | 91.2 |
| 42 | 0.979 | 961 | 4 | 43 | 1.2 | 3,234 | 92.4 |
| 43 | 1.063 | 962 | 4 | 50 | 1.4 | 3,284 | 93.9 |
| 44 | 1.149 | 963 | 4 | 62 | 1.8 | 3,346 | 95.6 |
| 45 | 1.237 | 964 | 4 | 36 | 1.0 | 3,382 | 96.7 |
| 46 | 1.329 | 966 | 4 | 22 | 0.6 | 3,404 | 97.3 |
| 47 | 1.424 | 967 | 4 | 37 | 1.1 | 3,441 | 98.3 |
| 48 | 1.523 | 968 | 4 | 17 | 0.5 | 3,458 | 98.8 |
| 49 | 1.627 | 970 | 5 | 16 | 0.5 | 3,474 | 99.3 |
| 50 | 1.736 | 971 | 5 | 9 | 0.3 | 3,483 | 99.5 |
| 51 | 1.852 | 973 | 5 | 3 | 0.1 | 3,486 | 99.6 |
| 52 | 1.976 | 975 | 5 | 6 | 0.2 | 3,492 | 99.8 |
| 53 | 2.110 | 977 | 5 | 4 | 0.1 | 3,496 | 99.9 |
| 54 | 2.256 | 979 | 5 | 2 | 0.1 | 3,498 | 100.0 |
| 58 | 3.059 | 990 | 7 | 1 | 0.0 | 3,499 | 100.0 |

### Appendix 7.B: Demographic Student Group Summaries of Participation

**Notes:**

* Data collected for Spanish instruction program types and percentage of daily instruction in Spanish are derived from the student survey as part of the operational assessment.
* Percent of valid scores for individual student demographic groups may not sum to 100 because of rounding.
* Data for the category of “Received in Spanish in the 2021–22 school year—Spanish as a Foreign Language Program” is available for grades six through eight and for high school.

Table 7.B. **Demographic Summary** for Grade **Three**

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| All students | 11,105 | 100.00 |
| Male | 5,480 | 49.35 |
| Female | 5,623 | 50.63 |
| Nonbinary | 2 | 0.02 |
| American Indian or Alaska Native | 24 | 0.22 |
| Asian | 75 | 0.68 |
| Native Hawaiian or Other Pacific Islander | 8 | 0.07 |
| Filipino | 72 | 0.65 |
| Hispanic or Latino | 9,700 | 87.35 |
| Black or African American | 231 | 2.08 |
| White | 714 | 6.43 |
| Two or more races | 281 | 2.53 |
| Unknown | 0 | 0.00 |
| English only | 3,921 | 35.31 |
| IFEP | 949 | 8.55 |
| EL | 5,073 | 45.68 |
| RFEP | 1,160 | 10.45 |
| Ever-ELs (EL or RFEP) | 6,233 | 56.13 |
| To be determined | 1 | 0.01 |
| English proficiency unknown | 1 | 0.01 |
| No special education services | 10,332 | 93.04 |
| Special education services | 773 | 6.96 |
| Economically disadvantaged | 7,899 | 71.13 |
| Not economically disadvantaged | 3,206 | 28.87 |

Table 7.B.1 *(continuation)*

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| In US schools less than 12 months | 410 | 3.69 |
| In US schools 12 months or more | 10,695 | 96.31 |
| Received Instruction in Spanish in the 2021–22 School Year—Total | 10,573 | 95.21 |
| Received Instruction in Spanish in the 2021–22 School Year—One-Way Immersion Program | 549 | 4.94 |
| Received Instruction in Spanish in the 2021–22 School Year—Dual-Language Immersion Program | 8,964 | 80.72 |
| Received Instruction in Spanish in the 2021–22 School Year—Developmental Bilingual Program | 616 | 5.55 |
| Received Instruction in Spanish in the 2021–22 School Year—Heritage Language or Indigenous Language Program | 146 | 1.31 |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 299 | 2.69 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 3,936 | 35.44 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 5,066 | 45.62 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 1,272 | 11.45 |

Table 7.B. **Demographic Summary** for Grade **Four**

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| All students | 10,119 | 100.00 |
| Male | 4,951 | 48.93 |
| Female | 5,168 | 51.07 |
| Nonbinary | 0 | 0.00 |
| American Indian or Alaska Native | 19 | 0.19 |
| Asian | 77 | 0.76 |
| Native Hawaiian or Other Pacific Islander | 4 | 0.04 |
| Filipino | 83 | 0.82 |
| Hispanic or Latino | 8,814 | 87.10 |
| Black or African American | 146 | 1.44 |
| White | 710 | 7.02 |
| Two or more races | 266 | 2.63 |
| Unknown | 0 | 0.00 |
| English only | 3,400 | 33.60 |
| IFEP | 520 | 5.14 |
| EL | 4,404 | 43.52 |
| RFEP | 1,790 | 17.69 |
| Ever-ELs (EL or RFEP) | 6,194 | 61.21 |
| To be determined | 4 | 0.04 |
| English proficiency unknown | 1 | 0.01 |
| No special education services | 9,452 | 93.41 |
| Special education services | 667 | 6.59 |
| Economically disadvantaged | 6,990 | 69.08 |
| Not economically disadvantaged | 3,129 | 30.92 |
| In US schools less than 12 months | 369 | 3.65 |
| In US schools 12 months or more | 9,750 | 96.35 |
| Received Instruction in Spanish in the 2021–22 School Year—Total | 9,739 | 96.24 |
| Received Instruction in Spanish in the 2021–22 School Year—One-Way Immersion Program | 296 | 2.93 |
| Received Instruction in Spanish in the 2021–22 School Year—Dual-Language Immersion Program | 8,597 | 84.96 |
| Received Instruction in Spanish in the 2021–22 School Year—Developmental Bilingual Program | 432 | 4.27 |
| Received Instruction in Spanish in the 2021–22 School Year—Heritage Language or Indigenous Language Program | 110 | 1.09 |

Table 7.B.2 *(continuation)*

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 292 | 2.89 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 5,128 | 50.68 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 3,467 | 34.26 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 852 | 8.42 |

Table 7.B. **Demographic Summary** for Grade **Five**

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| All students | 8,434 | 100.00 |
| Male | 4,092 | 48.52 |
| Female | 4,340 | 51.46 |
| Nonbinary | 2 | 0.02 |
| American Indian or Alaska Native | 14 | 0.17 |
| Asian | 67 | 0.79 |
| Native Hawaiian or Other Pacific Islander | 6 | 0.07 |
| Filipino | 62 | 0.74 |
| Hispanic or Latino | 7,354 | 87.19 |
| Black or African American | 146 | 1.73 |
| White | 583 | 6.91 |
| Two or more races | 202 | 2.40 |
| Unknown | 0 | 0.00 |
| English only | 2,614 | 30.99 |
| IFEP | 423 | 5.02 |
| EL | 3,110 | 36.87 |
| RFEP | 2,284 | 27.08 |
| Ever-ELs (EL or RFEP) | 5,394 | 63.96 |
| To be determined | 3 | 0.04 |
| English proficiency unknown | 0 | 0.00 |
| No special education services | 7,888 | 93.53 |
| Special education services | 546 | 6.47 |
| Economically disadvantaged | 5,793 | 68.69 |
| Not economically disadvantaged | 2,641 | 31.31 |
| In US schools less than 12 months | 343 | 4.07 |
| In US schools 12 months or more | 8,091 | 95.93 |
| Received Instruction in Spanish in the 2021–22 School Year—Total | 8,167 | 96.83 |
| Received Instruction in Spanish in the 2021–22 School Year—One-Way Immersion Program | 254 | 3.01 |
| Received Instruction in Spanish in the 2021–22 School Year—Dual-Language Immersion Program | 7,283 | 86.35 |
| Received Instruction in Spanish in the 2021–22 School Year—Developmental Bilingual Program | 338 | 4.01 |
| Received Instruction in Spanish in the 2021–22 School Year—Heritage Language or Indigenous Language Program | 52 | 0.62 |

Table 7.B.3 *(continuation)*

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 313 | 3.71 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 5,261 | 62.38 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 1,969 | 23.35 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 624 | 7.40 |

Table 7.B. **Demographic Summary** for Grade **Six**

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| All students | 5,892 | 100.00 |
| Male | 2,835 | 48.12 |
| Female | 3,056 | 51.87 |
| Nonbinary | 1 | 0.02 |
| American Indian or Alaska Native | 10 | 0.17 |
| Asian | 54 | 0.92 |
| Native Hawaiian or Other Pacific Islander | 5 | 0.08 |
| Filipino | 46 | 0.78 |
| Hispanic or Latino | 5,155 | 87.49 |
| Black or African American | 86 | 1.46 |
| White | 379 | 6.43 |
| Two or more races | 157 | 2.66 |
| Unknown | 0 | 0.00 |
| English only | 1,809 | 30.70 |
| IFEP | 318 | 5.40 |
| EL | 1,868 | 31.70 |
| RFEP | 1,890 | 32.08 |
| Ever-ELs (EL or RFEP) | 3,758 | 63.78 |
| To be determined | 6 | 0.10 |
| English proficiency unknown | 1 | 0.02 |
| No special education services | 5,556 | 94.30 |
| Special education services | 336 | 5.70 |
| Economically disadvantaged | 3,789 | 64.31 |
| Not economically disadvantaged | 2,103 | 35.69 |
| In US schools less than 12 months | 277 | 4.70 |
| In US schools 12 months or more | 5,615 | 95.30 |
| Received Instruction in Spanish in the 2021–22 School Year—Total | 5,729 | 97.23 |
| Received Instruction in Spanish in the 2021–22 School Year—One-Way Immersion Program | 67 | 1.14 |
| Received Instruction in Spanish in the 2021–22 School Year—Dual-Language Immersion Program | 5,346 | 90.73 |
| Received Instruction in Spanish in the 2021–22 School Year—Developmental Bilingual Program | 155 | 2.63 |
| Received Instruction in Spanish in the 2021–22 School Year—Heritage Language or Indigenous Language Program | 26 | 0.44 |
| Received Instruction in Spanish in the 2021–22 School Year—Spanish as a Foreign Language Program | 36 | 0.61 |

Table 7.B.4 *(continuation)*

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 377 | 6.40 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 3,691 | 62.64 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 1,250 | 21.22 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 411 | 6.98 |

Table 7.B. **Demographic Summary** for Grade **Seven**

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| All students | 3,833 | 100.00 |
| Male | 1,830 | 47.74 |
| Female | 2,001 | 52.20 |
| Nonbinary | 2 | 0.05 |
| American Indian or Alaska Native | 7 | 0.18 |
| Asian | 25 | 0.65 |
| Native Hawaiian or Other Pacific Islander | 1 | 0.03 |
| Filipino | 9 | 0.23 |
| Hispanic or Latino | 3,341 | 87.16 |
| Black or African American | 37 | 0.97 |
| White | 291 | 7.59 |
| Two or more races | 122 | 3.18 |
| Unknown | 0 | 0.00 |
| English only | 1,021 | 26.64 |
| IFEP | 179 | 4.67 |
| EL | 1,119 | 29.19 |
| RFEP | 1,508 | 39.34 |
| Ever-ELs (EL or RFEP) | 2,627 | 68.54 |
| To be determined | 5 | 0.13 |
| English proficiency unknown | 1 | 0.03 |
| No special education services | 3,650 | 95.23 |
| Special education services | 183 | 4.77 |
| Economically disadvantaged | 2,572 | 67.10 |
| Not economically disadvantaged | 1,261 | 32.90 |
| In US schools less than 12 months | 280 | 7.30 |
| In US schools 12 months or more | 3,553 | 92.70 |
| Received Instruction in Spanish in the 2021–22 School Year—Total | 3,687 | 96.19 |
| Received Instruction in Spanish in the 2021–22 School Year—One-Way Immersion Program | 68 | 1.77 |
| Received Instruction in Spanish in the 2021–22 School Year—Dual-Language Immersion Program | 3,241 | 84.56 |
| Received Instruction in Spanish in the 2021–22 School Year—Developmental Bilingual Program | 157 | 4.10 |
| Received Instruction in Spanish in the 2021–22 School Year—Heritage Language or Indigenous Language Program | 48 | 1.25 |
| Received Instruction in Spanish in the 2021–22 School Year—Spanish as a Foreign Language Program | 72 | 1.88 |

Table 7.B.5 *(continuation)*

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 623 | 16.25 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 2,312 | 60.32 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 477 | 12.44 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 275 | 7.17 |

Table 7.B. **Demographic Summary** for Grade **Eight**

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| All students | 3,239 | 100.00 |
| Male | 1,542 | 47.61 |
| Female | 1,697 | 52.39 |
| Nonbinary | 0 | 0.00 |
| American Indian or Alaska Native | 4 | 0.12 |
| Asian | 14 | 0.43 |
| Native Hawaiian or Other Pacific Islander | 3 | 0.09 |
| Filipino | 11 | 0.34 |
| Hispanic or Latino | 2,852 | 88.05 |
| Black or African American | 38 | 1.17 |
| White | 238 | 7.35 |
| Two or more races | 79 | 2.44 |
| Unknown | 0 | 0.00 |
| English only | 780 | 24.08 |
| IFEP | 111 | 3.43 |
| EL | 886 | 27.35 |
| RFEP | 1,459 | 45.04 |
| Ever-ELs (EL or RFEP) | 2,345 | 72.40 |
| To be determined | 3 | 0.09 |
| English proficiency unknown | 0 | 0.00 |
| No special education services | 3,090 | 95.40 |
| Special education services | 149 | 4.60 |
| Economically disadvantaged | 2,255 | 69.62 |
| Not economically disadvantaged | 984 | 30.38 |
| In US schools less than 12 months | 272 | 8.40 |
| In US schools 12 months or more | 2,967 | 91.60 |
| Received Instruction in Spanish in the 2021–22 School Year—Total | 3,078 | 95.03 |
| Received Instruction in Spanish in the 2021–22 School Year—One-Way Immersion Program | 57 | 1.76 |
| Received Instruction in Spanish in the 2021–22 School Year—Dual-Language Immersion Program | 2,730 | 84.29 |
| Received Instruction in Spanish in the 2021–22 School Year—Developmental Bilingual Program | 107 | 3.30 |
| Received Instruction in Spanish in the 2021–22 School Year—Heritage Language or Indigenous Language Program | 29 | 0.90 |
| Received Instruction in Spanish in the 2021–22 School Year—Spanish as a Foreign Language Program | 48 | 1.48 |

Table 7.B.6 *(continuation)*

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 825 | 25.47 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 1,618 | 49.95 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 363 | 11.21 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 272 | 8.40 |

Table 7.B. **Demographic** Student Group **Summary** for High **School**

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| All students | 3,513 | 100.00 |
| Male | 1,709 | 48.65 |
| Female | 1,804 | 51.35 |
| Nonbinary | 0 | 0.00 |
| American Indian or Alaska Native | 3 | 0.09 |
| Asian | 5 | 0.14 |
| Native Hawaiian or Other Pacific Islander | 0 | 0.00 |
| Filipino | 4 | 0.11 |
| Hispanic or Latino | 3,188 | 90.75 |
| Black or African American | 39 | 1.11 |
| White | 92 | 2.62 |
| Two or more races | 182 | 5.18 |
| Unknown | 0 | 0.00 |
| English only | 464 | 13.21 |
| IFEP | 104 | 2.96 |
| EL | 1,387 | 39.48 |
| RFEP | 1,556 | 44.29 |
| ADEL | 0 | 0.00 |
| Ever-ELs (EL, RFEP, or ADEL) | 2,943 | 83.77 |
| To be determined | 2 | 0.06 |
| English proficiency unknown | 0 | 0.00 |
| No special education services | 3,392 | 96.56 |
| Special education services | 121 | 3.44 |
| Economically disadvantaged | 2,656 | 75.60 |
| Not economically disadvantaged | 857 | 24.40 |
| In US schools less than 12 months | 785 | 22.35 |
| In US schools 12 months or more | 2,728 | 77.65 |
| Received Instruction in Spanish in the 2021–22 School Year—Total | 3,087 | 87.87 |
| Received Instruction in Spanish in the 2021–22 School Year—One-Way Immersion Program | 196 | 5.58 |
| Received Instruction in Spanish in the 2021–22 School Year—Dual-Language Immersion Program | 802 | 22.83 |
| Received Instruction in Spanish in the 2021–22 School Year—Developmental Bilingual Program | 137 | 3.90 |
| Received Instruction in Spanish in the 2021–22 School Year—Heritage Language or Indigenous Language Program | 473 | 13.46 |
| Received Instruction in Spanish in the 2021–22 School Year—Spanish as a Foreign Language Program | 1,217 | 34.64 |

Table 7.B.7 *(continuation)*

|  |  |  |
| --- | --- | --- |
| **Student Group** | **Number of Valid Scores** | **Percent of Valid Scores** |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 1,249 | 35.55 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 841 | 23.94 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 451 | 12.84 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 546 | 15.54 |

### Appendix 7.C: Demographic Student Group Summaries of Performance

**Notes:**

* To protect privacy when the number of students in a student group is 10 or fewer, the summary statistics at the test and reporting levels are not reported and are presented as “N/A” in the tables in appendix 7.C.
* Percentages in these tables may not sum up to 100 because of rounding.
* Table 7.C.1 through table 7.C.7, include students who have completed testing with a valid reporting scale score.
* Data for the category of “Received in Spanish in the 2021–22 school year—Spanish as a Foreign Language Program” is available for grades six through eight and for high school.
* SD = standard deviation.

Table 7.C. Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for Grade Three

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| All students | 11,105 | 344 | 10.3 | 71.9 | 18.7 | 9.3 |
| Male | 5,480 | 343 | 10.1 | 75.5 | 16.2 | 8.2 |
| Female | 5,623 | 345 | 10.5 | 68.4 | 21.2 | 10.4 |
| Nonbinary | 2 | N/A | N/A | N/A | N/A | N/A |
| American Indian or Alaska Native | 24 | 342 | 10.5 | 79.2 | 8.3 | 12.5 |
| Asian | 75 | 348 | 12.5 | 56.0 | 25.3 | 18.7 |
| Native Hawaiian or Other Pacific Islander | 8 | N/A | N/A | N/A | N/A | N/A |
| Filipino | 72 | 347 | 9.6 | 59.7 | 30.6 | 9.7 |
| Hispanic or Latino | 9,700 | 344 | 10.2 | 72.2 | 18.7 | 9.1 |
| Black or African American | 231 | 338 | 7.6 | 92.6 | 4.3 | 3.0 |
| White | 714 | 346 | 11.2 | 66.7 | 21.4 | 11.9 |
| Two or more races | 281 | 346 | 11.3 | 63.7 | 21.4 | 14.9 |
| Unknown | 0 | N/A | N/A | N/A | N/A | N/A |

Table 7.C.1 *(continuation one)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| English only | 3,921 | 343 | 10.2 | 75.2 | 16.5 | 8.3 |
| Initial fluent English proficient (IFEP) | 949 | 348 | 11.6 | 55.7 | 26.2 | 18.0 |
| EL | 5,073 | 342 | 9.1 | 77.8 | 16.3 | 5.8 |
| Reclassified fluent English proficient (RFEP) | 1,160 | 350 | 11.3 | 48.2 | 30.9 | 20.9 |
| Ever-ELs (EL or RFEP) | 6,233 | 344 | 10.0 | 72.3 | 19.0 | 8.6 |
| To be determined | 1 | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 1 | N/A | N/A | N/A | N/A | N/A |
| No special education services | 10,332 | 344 | 10.4 | 70.8 | 19.4 | 9.8 |
| Special education services | 773 | 339 | 8.3 | 87.6 | 9.6 | 2.8 |
| Economically disadvantaged | 7,899 | 343 | 9.8 | 75.4 | 17.3 | 7.3 |
| Not economically disadvantaged | 3,206 | 346 | 11.2 | 63.3 | 22.3 | 14.4 |
| In US schools less than 12 months | 410 | 343 | 9.8 | 74.4 | 16.6 | 9.0 |
| In US schools 12 months or more | 10,695 | 344 | 10.3 | 71.8 | 18.8 | 9.3 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Total | 10,573 | 344 | 10.4 | 70.9 | 19.4 | 9.7 |
| Received Instruction in Spanish in the 2021–‍22 School Year—One-Way Immersion Program | 549 | 341 | 9.4 | 82.5 | 11.7 | 5.8 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Dual-Language Immersion Program | 8,964 | 345 | 10.6 | 69.0 | 20.5 | 10.5 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Developmental Bilingual Program | 616 | 342 | 8.6 | 80.2 | 15.6 | 4.2 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Heritage Language or Indigenous Language Program | 146 | 338 | 6.4 | 95.2 | 4.1 | 0.7 |

Table 7.C.1 *(continuation two)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 299 | 339 | 7.4 | 90.6 | 6.4 | 3.0 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 3,936 | 343 | 10.0 | 75.4 | 16.9 | 7.6 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 5,066 | 345 | 10.7 | 65.8 | 22.2 | 11.9 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 1,272 | 344 | 9.9 | 72.8 | 18.6 | 8.6 |

Table 7.C. Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for Grade Four

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| All students | 10,119 | 446 | 9.4 | 65.4 | 25.2 | 9.4 |
| Male | 4,951 | 445 | 9.2 | 69.0 | 22.9 | 8.2 |
| Female | 5,168 | 446 | 9.5 | 62.0 | 27.4 | 10.6 |
| Nonbinary | 0 | N/A | N/A | N/A | N/A | N/A |
| American Indian or Alaska Native | 19 | 442 | 7.9 | 84.2 | 10.5 | 5.3 |
| Asian | 77 | 449 | 11.7 | 51.9 | 33.8 | 14.3 |
| Native Hawaiian or Other Pacific Islander | 4 | N/A | N/A | N/A | N/A | N/A |
| Filipino | 83 | 451 | 9.6 | 43.4 | 39.8 | 16.9 |
| Hispanic or Latino | 8,814 | 446 | 9.3 | 66.2 | 24.9 | 9.0 |
| Black or African American | 146 | 441 | 8.0 | 82.2 | 15.1 | 2.7 |
| White | 710 | 448 | 10.3 | 57.3 | 29.2 | 13.5 |
| Two or more races | 266 | 447 | 9.8 | 60.5 | 25.6 | 13.9 |
| Unknown | 0 | N/A | N/A | N/A | N/A | N/A |
| English only | 3,400 | 445 | 9.4 | 68.2 | 23.4 | 8.3 |
| IFEP | 520 | 451 | 10.7 | 44.6 | 34.4 | 21.0 |
| EL | 4,404 | 444 | 8.0 | 75.2 | 20.3 | 4.5 |
| RFEP | 1,790 | 451 | 9.9 | 41.9 | 37.9 | 20.2 |
| Ever-ELs (EL or RFEP) | 6,194 | 446 | 9.2 | 65.6 | 25.4 | 9.0 |
| To be determined | 4 | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 1 | N/A | N/A | N/A | N/A | N/A |
| No special education services | 9,452 | 446 | 9.4 | 63.9 | 26.2 | 9.8 |
| Special education services | 667 | 441 | 7.7 | 86.1 | 10.3 | 3.6 |
| Economically disadvantaged | 6,990 | 445 | 9.1 | 69.2 | 23.0 | 7.7 |
| Not economically disadvantaged | 3,129 | 448 | 9.9 | 56.8 | 30.0 | 13.2 |

Table 7.C.2 *(continuation)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| In US schools less than 12 months | 369 | 444 | 8.6 | 71.5 | 22.5 | 6.0 |
| In US schools 12 months or more | 9,750 | 446 | 9.4 | 65.2 | 25.3 | 9.5 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Total | 9,739 | 446 | 9.4 | 64.8 | 25.6 | 9.6 |
| Received Instruction in Spanish in the 2021–‍22 School Year—One-Way Immersion Program | 296 | 442 | 8.5 | 84.5 | 11.5 | 4.1 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Dual-Language Immersion Program | 8,597 | 446 | 9.5 | 63.0 | 26.8 | 10.2 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Developmental Bilingual Program | 432 | 444 | 9.1 | 73.8 | 19.9 | 6.3 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Heritage Language or Indigenous Language Program | 110 | 439 | 5.6 | 94.5 | 4.5 | 0.9 |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 292 | 443 | 8.8 | 76.0 | 20.2 | 3.8 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 5,128 | 446 | 9.5 | 65.8 | 24.8 | 9.3 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 3,467 | 447 | 9.4 | 61.8 | 27.3 | 10.9 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 852 | 445 | 9.1 | 67.1 | 24.9 | 8.0 |

Table 7.C. Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for Grade Five

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| All students | 8,434 | 546 | 10.0 | 51.5 | 37.9 | 10.7 |
| Male | 4,092 | 545 | 9.7 | 57.0 | 34.3 | 8.7 |
| Female | 4,340 | 547 | 10.2 | 46.2 | 41.3 | 12.5 |
| Nonbinary | 2 | N/A | N/A | N/A | N/A | N/A |
| American Indian or Alaska Native | 14 | 547 | 9.2 | 42.9 | 42.9 | 14.3 |
| Asian | 67 | 548 | 10.8 | 40.3 | 44.8 | 14.9 |
| Native Hawaiian or Other Pacific Islander | 6 | N/A | N/A | N/A | N/A | N/A |
| Filipino | 62 | 551 | 11.0 | 29.0 | 53.2 | 17.7 |
| Hispanic or Latino | 7,354 | 546 | 9.9 | 51.9 | 37.8 | 10.3 |
| Black or African American | 146 | 542 | 8.3 | 70.5 | 26.7 | 2.7 |
| White | 583 | 547 | 10.8 | 47.9 | 37.4 | 14.8 |
| Two or more races | 202 | 548 | 10.9 | 44.1 | 42.1 | 13.9 |
| Unknown | 0 | N/A | N/A | N/A | N/A | N/A |
| English only | 2,614 | 545 | 9.9 | 57.0 | 34.4 | 8.5 |
| IFEP | 423 | 551 | 10.8 | 30.0 | 46.8 | 23.2 |
| EL | 3,110 | 543 | 8.6 | 65.0 | 30.5 | 4.5 |
| RFEP | 2,284 | 551 | 9.9 | 30.6 | 50.2 | 19.2 |
| Ever-ELs (EL or RFEP) | 5,394 | 546 | 9.9 | 50.4 | 38.8 | 10.7 |
| To be determined | 3 | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 0 | N/A | N/A | N/A | N/A | N/A |
| No special education services | 7,888 | 547 | 10.0 | 49.4 | 39.3 | 11.2 |
| Special education services | 546 | 540 | 8.3 | 80.8 | 16.7 | 2.6 |
| Economically disadvantaged | 5,793 | 545 | 9.7 | 54.8 | 36.1 | 9.2 |
| Not economically disadvantaged | 2,641 | 548 | 10.4 | 44.2 | 41.8 | 14.0 |

Table 7.C.3 *(continuation)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| In US schools less than 12 months | 343 | 546 | 10.1 | 54.5 | 34.1 | 11.4 |
| In US schools 12 months or more | 8,091 | 546 | 10.0 | 51.3 | 38.0 | 10.6 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Total | 8,167 | 546 | 10.0 | 50.9 | 38.2 | 10.9 |
| Received Instruction in Spanish in the 2021–‍22 School Year—One-Way Immersion Program | 254 | 543 | 9.3 | 67.7 | 26.0 | 6.3 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Dual-Language Immersion Program | 7,283 | 547 | 10.0 | 49.1 | 39.5 | 11.4 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Developmental Bilingual Program | 338 | 543 | 9.7 | 65.7 | 27.8 | 6.5 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Heritage Language or Indigenous Language Program | 52 | 539 | 8.2 | 82.7 | 13.5 | 3.8 |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 313 | 543 | 9.7 | 63.9 | 27.5 | 8.6 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 5,261 | 547 | 10.1 | 48.3 | 39.9 | 11.8 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 1,969 | 546 | 9.7 | 52.6 | 37.9 | 9.5 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 624 | 544 | 9.6 | 61.5 | 30.3 | 8.2 |

Table 7.C. Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for Grade Six

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| All students | 5,892 | 650 | 10.6 | 46.3 | 35.2 | 18.5 |
| Male | 2,835 | 648 | 10.6 | 51.8 | 33.0 | 15.2 |
| Female | 3,056 | 651 | 10.5 | 41.1 | 37.3 | 21.5 |
| Nonbinary | 1 | N/A | N/A | N/A | N/A | N/A |
| American Indian or Alaska Native | 10 | N/A | N/A | N/A | N/A | N/A |
| Asian | 54 | 654 | 12.5 | 31.5 | 35.2 | 33.3 |
| Native Hawaiian or Other Pacific Islander | 5 | N/A | N/A | N/A | N/A | N/A |
| Filipino | 46 | 653 | 9.4 | 28.3 | 47.8 | 23.9 |
| Hispanic or Latino | 5,155 | 649 | 10.6 | 46.7 | 35.4 | 17.9 |
| Black or African American | 86 | 645 | 9.8 | 64.0 | 23.3 | 12.8 |
| White | 379 | 652 | 10.7 | 39.1 | 35.1 | 25.9 |
| Two or more races | 157 | 648 | 11.2 | 50.3 | 33.1 | 16.6 |
| Unknown | 0 | N/A | N/A | N/A | N/A | N/A |
| English only | 1,809 | 649 | 10.5 | 46.4 | 35.0 | 18.5 |
| IFEP | 318 | 657 | 11.5 | 23.3 | 34.6 | 42.1 |
| EL | 1,868 | 644 | 8.8 | 67.9 | 26.6 | 5.6 |
| RFEP | 1,890 | 653 | 10.0 | 28.5 | 44.2 | 27.4 |
| Ever-ELs (EL or RFEP) | 3,758 | 649 | 10.4 | 48.1 | 35.4 | 16.5 |
| To be determined | 6 | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 1 | N/A | N/A | N/A | N/A | N/A |
| No special education services | 5,556 | 650 | 10.6 | 44.1 | 36.4 | 19.4 |
| Special education services | 336 | 641 | 7.8 | 81.8 | 15.2 | 3.0 |
| Economically disadvantaged | 3,789 | 648 | 10.3 | 50.5 | 34.3 | 15.2 |
| Not economically disadvantaged | 2,103 | 651 | 11.0 | 38.6 | 36.9 | 24.5 |

Table 7.C.4 *(continuation)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| In US schools less than 12 months | 277 | 646 | 10.1 | 63.2 | 26.0 | 10.8 |
| In US schools 12 months or more | 5,615 | 650 | 10.6 | 45.4 | 35.7 | 18.9 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Total | 5,729 | 650 | 10.6 | 45.5 | 35.6 | 18.9 |
| Received Instruction in Spanish in the 2021–‍22 School Year—One-Way Immersion Program | 67 | 642 | 9.2 | 73.1 | 23.9 | 3.0 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Dual-Language Immersion Program | 5,346 | 650 | 10.6 | 43.8 | 36.4 | 19.9 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Developmental Bilingual Program | 155 | 645 | 9.1 | 63.2 | 29.7 | 7.1 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Heritage Language or Indigenous Language Program | 26 | 643 | 10.8 | 76.9 | 15.4 | 7.7 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Spanish as a Foreign Language Program | 36 | 640 | 7.8 | 88.9 | 8.3 | 2.8 |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 377 | 650 | 11.0 | 46.9 | 32.1 | 21.0 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 3,691 | 650 | 10.7 | 43.5 | 36.1 | 20.3 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 1,250 | 649 | 10.4 | 46.1 | 36.0 | 17.9 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 411 | 646 | 9.1 | 59.6 | 33.1 | 7.3 |

Table 7.C. Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for Grade Seven

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| All students | 3,833 | 741 | 10.1 | 61.8 | 32.6 | 5.6 |
| Male | 1,830 | 740 | 9.7 | 68.7 | 27.2 | 4.2 |
| Female | 2,001 | 743 | 10.3 | 55.5 | 37.6 | 6.9 |
| Nonbinary | 2 | N/A | N/A | N/A | N/A | N/A |
| American Indian or Alaska Native | 7 | N/A | N/A | N/A | N/A | N/A |
| Asian | 25 | 745 | 13.3 | 44.0 | 44.0 | 12.0 |
| Native Hawaiian or Other Pacific Islander | 1 | N/A | N/A | N/A | N/A | N/A |
| Filipino | 9 | N/A | N/A | N/A | N/A | N/A |
| Hispanic or Latino | 3,341 | 741 | 10.0 | 62.6 | 32.4 | 5.1 |
| Black or African American | 37 | 738 | 8.4 | 75.7 | 24.3 | 0.0 |
| White | 291 | 744 | 11.2 | 50.9 | 40.2 | 8.9 |
| Two or more races | 122 | 742 | 11.1 | 63.9 | 23.8 | 12.3 |
| Unknown | 0 | N/A | N/A | N/A | N/A | N/A |
| English only | 1,021 | 741 | 10.4 | 60.8 | 33.4 | 5.8 |
| IFEP | 179 | 746 | 11.7 | 43.0 | 42.5 | 14.5 |
| EL | 1,119 | 737 | 8.4 | 79.3 | 19.2 | 1.5 |
| RFEP | 1,508 | 744 | 9.9 | 51.7 | 40.9 | 7.4 |
| Ever-ELs (EL or RFEP) | 2,627 | 741 | 9.8 | 63.5 | 31.7 | 4.9 |
| To be determined | 5 | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 1 | N/A | N/A | N/A | N/A | N/A |
| No special education services | 3,650 | 742 | 10.1 | 60.3 | 33.9 | 5.8 |
| Special education services | 183 | 734 | 8.3 | 90.7 | 8.2 | 1.1 |
| Economically disadvantaged | 2,572 | 740 | 9.8 | 65.4 | 30.3 | 4.4 |
| Not economically disadvantaged | 1,261 | 743 | 10.5 | 54.5 | 37.4 | 8.1 |

Table 7.C.5 *(continuation)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| In US schools less than 12 months | 280 | 739 | 8.9 | 71.4 | 25.4 | 3.2 |
| In US schools 12 months or more | 3,553 | 742 | 10.2 | 61.0 | 33.2 | 5.8 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Total | 3,687 | 742 | 10.1 | 60.9 | 33.3 | 5.8 |
| Received Instruction in Spanish in the 2021–‍22 School Year—One-Way Immersion Program | 68 | 739 | 9.6 | 70.6 | 25.0 | 4.4 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Dual-Language Immersion Program | 3,241 | 742 | 10.1 | 59.1 | 34.9 | 6.0 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Developmental Bilingual Program | 157 | 739 | 9.3 | 69.4 | 26.8 | 3.8 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Heritage Language or Indigenous Language Program | 48 | 735 | 13.5 | 79.2 | 12.5 | 8.3 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Spanish as a Foreign Language Program | 72 | 735 | 8.3 | 86.1 | 12.5 | 1.4 |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 623 | 741 | 10.3 | 64.2 | 30.3 | 5.5 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 2,312 | 742 | 10.4 | 57.1 | 36.2 | 6.8 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 477 | 740 | 9.2 | 67.7 | 29.6 | 2.7 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 275 | 739 | 8.3 | 74.5 | 21.8 | 3.6 |

Table 7.C. Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for Grade Eight

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| All students | 3,239 | 845 | 11.2 | 60.6 | 28.0 | 11.4 |
| Male | 1,542 | 843 | 11.0 | 68.7 | 22.6 | 8.7 |
| Female | 1,697 | 847 | 11.1 | 53.3 | 32.8 | 13.9 |
| Nonbinary | 0 | N/A | N/A | N/A | N/A | N/A |
| American Indian or Alaska Native | 4 | N/A | N/A | N/A | N/A | N/A |
| Asian | 14 | 855 | 14.6 | 28.6 | 42.9 | 28.6 |
| Native Hawaiian or Other Pacific Islander | 3 | N/A | N/A | N/A | N/A | N/A |
| Filipino | 11 | 855 | 12.7 | 18.2 | 54.5 | 27.3 |
| Hispanic or Latino | 2,852 | 845 | 11.0 | 61.9 | 27.5 | 10.7 |
| Black or African American | 38 | 843 | 9.4 | 68.4 | 21.1 | 10.5 |
| White | 238 | 849 | 12.1 | 47.5 | 34.0 | 18.5 |
| Two or more races | 79 | 843 | 12.2 | 63.3 | 24.1 | 12.7 |
| Unknown | 0 | N/A | N/A | N/A | N/A | N/A |
| English only | 780 | 846 | 11.7 | 56.9 | 29.9 | 13.2 |
| IFEP | 111 | 851 | 12.3 | 44.1 | 28.8 | 27.0 |
| EL | 886 | 840 | 9.2 | 81.4 | 15.5 | 3.2 |
| RFEP | 1,459 | 848 | 10.8 | 51.2 | 34.5 | 14.3 |
| Ever-ELs (EL or RFEP) | 2,345 | 845 | 10.9 | 62.6 | 27.3 | 10.1 |
| To be determined | 3 | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 0 | N/A | N/A | N/A | N/A | N/A |
| No special education services | 3,090 | 846 | 11.2 | 59.3 | 28.8 | 11.9 |
| Special education services | 149 | 838 | 8.6 | 87.9 | 10.1 | 2.0 |
| Economically disadvantaged | 2,255 | 845 | 10.8 | 63.3 | 26.9 | 9.8 |
| Not economically disadvantaged | 984 | 847 | 12.0 | 54.5 | 30.5 | 15.0 |

Table 7.C.6 *(continuation)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| In US schools less than 12 months | 272 | 842 | 10.6 | 71.0 | 21.7 | 7.4 |
| In US schools 12 months or more | 2,967 | 846 | 11.2 | 59.7 | 28.5 | 11.8 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Total | 3,078 | 846 | 11.2 | 59.7 | 28.5 | 11.8 |
| Received Instruction in Spanish in the 2021–‍22 School Year—One-Way Immersion Program | 57 | 839 | 9.3 | 84.2 | 12.3 | 3.5 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Dual-Language Immersion Program | 2,730 | 846 | 11.2 | 57.5 | 29.9 | 12.6 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Developmental Bilingual Program | 107 | 842 | 9.6 | 79.4 | 13.1 | 7.5 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Heritage Language or Indigenous Language Program | 29 | 845 | 11.5 | 58.6 | 27.6 | 13.8 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Spanish as a Foreign Language Program | 48 | 836 | 7.7 | 89.6 | 10.4 | 0.0 |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 825 | 845 | 11.4 | 59.0 | 28.8 | 12.1 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 1,618 | 847 | 11.3 | 55.3 | 30.3 | 14.3 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 363 | 844 | 10.2 | 66.4 | 26.7 | 6.9 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 272 | 841 | 8.9 | 78.7 | 19.1 | 2.2 |

Table 7.C. Percent of Students in Each Score Reporting Range for Total Scores by Demographic Student Group for High School, All Grade Levels

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| All students | 3,513 | 946 | 10.0 | 65.7 | 23.6 | 10.7 |
| Male | 1,709 | 944 | 9.7 | 73.5 | 18.1 | 8.4 |
| Female | 1,804 | 948 | 10.1 | 58.3 | 28.8 | 12.9 |
| Nonbinary | 0 | N/A | N/A | N/A | N/A | N/A |
| American Indian or Alaska Native | 3 | N/A | N/A | N/A | N/A | N/A |
| Asian | 5 | N/A | N/A | N/A | N/A | N/A |
| Native Hawaiian or Other Pacific Islander | 0 | N/A | N/A | N/A | N/A | N/A |
| Filipino | 4 | N/A | N/A | N/A | N/A | N/A |
| Hispanic or Latino | 3,188 | 946 | 10.1 | 64.9 | 24.0 | 11.1 |
| Black or African American | 39 | 937 | 5.9 | 97.4 | 2.6 | 0.0 |
| White | 92 | 947 | 10.9 | 54.3 | 33.7 | 12.0 |
| Two or more races | 182 | 944 | 8.9 | 78.6 | 15.4 | 6.0 |
| Unknown | 0 | N/A | N/A | N/A | N/A | N/A |
| English only | 464 | 942 | 9.6 | 77.4 | 17.7 | 5.0 |
| IFEP | 104 | 949 | 11.4 | 56.7 | 19.2 | 24.0 |
| EL | 1,387 | 944 | 9.4 | 73.1 | 19.2 | 7.7 |
| RFEP | 1,556 | 948 | 10.1 | 56.2 | 29.6 | 14.2 |
| Adult English Learner (ADEL) | 0 | N/A | N/A | N/A | N/A | N/A |
| Ever-ELs (EL, RFEP, or ADEL) | 2,943 | 946 | 9.9 | 64.2 | 24.7 | 11.1 |
| To be determined | 2 | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 0 | N/A | N/A | N/A | N/A | N/A |
| No special education services | 3,392 | 946 | 10.0 | 64.7 | 24.3 | 11.0 |
| Special education services | 121 | 939 | 7.5 | 93.4 | 5.0 | 1.7 |
| Economically disadvantaged | 2,656 | 946 | 10.1 | 65.7 | 23.8 | 10.4 |
| Not economically disadvantaged | 857 | 946 | 9.9 | 65.6 | 22.9 | 11.6 |

Table 7.C.7 *(continuation)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number Valid Score** | **Scale Score Mean** | **Scale Score SD** | **Reporting Score Range 1** | **Reporting Score Range 2** | **Reporting Score Range 3** |
| In US schools less than 12 months | 785 | 946 | 9.6 | 68.3 | 22.0 | 9.7 |
| In US schools 12 months or more | 2,728 | 946 | 10.2 | 65.0 | 24.0 | 11.0 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Total | 3,087 | 946 | 9.9 | 66.1 | 23.6 | 10.3 |
| Received Instruction in Spanish in the 2021–‍22 School Year—One-Way Immersion Program | 196 | 945 | 9.4 | 70.4 | 23.5 | 6.1 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Dual-Language Immersion Program | 802 | 948 | 10.1 | 58.0 | 28.6 | 13.5 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Developmental Bilingual Program | 137 | 944 | 8.8 | 78.8 | 13.9 | 7.3 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Heritage Language or Indigenous Language Program | 473 | 947 | 10.0 | 58.6 | 28.1 | 13.3 |
| Received Instruction in Spanish in the 2021–‍22 School Year—Spanish as a Foreign Language Program | 1,217 | 944 | 9.9 | 70.7 | 20.8 | 8.5 |
| Percentage of School-Day Instruction Provided in Spanish—0–25% | 1,249 | 945 | 10.2 | 65.8 | 23.5 | 10.6 |
| Percentage of School-Day Instruction Provided in Spanish—26–50% | 841 | 947 | 10.2 | 60.0 | 26.2 | 13.8 |
| Percentage of School-Day Instruction Provided in Spanish—51–75% | 451 | 945 | 9.2 | 70.7 | 22.8 | 6.4 |
| Percentage of School-Day Instruction Provided in Spanish—76–100% | 546 | 945 | 9.1 | 72.3 | 20.5 | 7.1 |

## Psychometric Analyses

This chapter describes in detail the psychometric analyses conducted for the 2021–22 California Spanish Assessment (CSA) and the results of the analyses. These include classical item analyses and differential item functioning (DIF) analyses; item response theory (IRT) calibration; postequating for grades three through eight and high school; field test scaling; omission, expiration, and completion analyses; overall testing summaries as the means and standard deviations (SDs) of scale scores and theta values; the percentages of students at each score reporting range for each test; testing time analysis; reliability analysis; consistency and accuracy of the score reporting range classifications; and student group reliability.

### Overview

This chapter describes the psychometric analyses conducted by ETS for the CSA, including classical item analyses, DIF analyses, IRT calibration, and response time analyses, as well as analyses to support reliability and validity evidence.

#### Summary of the Analyses

The following list identifies the analyses that are conducted for a typical CSA administration. Each analysis is described in the subsequent narrative.

1. **Classical Item Analyses—**Classical item analyses for the CSA are discussed in section [*8.2 Classical Item Analyses*](#_Classical_Item_Analyses).
2. **DIF Analyses—**DIF analyses for the CSA are described in section [*8.3 Differential Item Functioning Analyses*](#_Differential_Item_Functioning).
3. **IRT Analyses—**IRT analyses, including calibration, postequating, and scaling for the CSA, are described in section [*8.4 Item Response Theory Analyses*](#_Item_Response_Theory).
4. **Testing Time Analyses—**Response time analyses are described in section [*8.5 Testing Time Analyses*](#_Testing_Time_Analyses).
5. **Reliability Analyses—**Reliability estimation for the CSA is illustrated in section [*8.6 Reliability Analyses*](#_Reliability_Analyses).
6. **Validity Evidence—**Validity evidence related to the CSA is discussed in section [*8.7 Validity Evidence*](#_Validity_Evidence).

#### Samples Used for Analyses

In general, analyses included in this technical report are based on all valid scores in the tested population. The actual data sample used depends on the time that data source became available as well as the information contained in the data to meet the analysis timeline.

Both classical item analysis and IRT calibration include students who logged on to the test and answered at least one item. Table 8.1 summarizes the sample sizes of the data file available for different analyses. The IRT calibration analyses in [appendix 8.B](#_Appendix_8.B:_Item) were run using the version 1 of the production data file (“P1”) that was updated in June 2022. The classical item analyses in [appendix 8.A](#_Appendix_8.A:_Classical) and item-level DIF analyses in table 8.10 were based on complete data after the test window was closed in July 2022. All other analyses, such as the testing time analyses and reliability analyses, used the final version of the production data file for student reporting, which became available in August 2022. All data sources include all valid student scores.

Table 8.1 shows small differences in student counts among the data sources used for IRT calibration analysis, classical item analysis and DIF analysis, and the final production data file. A small number of student scores were excluded from the final production data as a result of the data validation process.

Table 8.1 Sample Size of Data Files Used for Different Analyses

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Calibration Sample N** | **Item Analysis and DIF Sample N** | **Final Production Data N** |
| Grade 3 | 9,753 | 9,828 | 11,105 |
| Grade 4 | 9,133 | 9,161 | 10,119 |
| Grade 5 | 7,725 | 7,760 | 8,434 |
| Grade 6 | 5,367 | 5,418 | 5,892 |
| Grade 7 | 3,684 | 3,728 | 3,833 |
| Grade 8 | 3,119 | 3,156 | 3,239 |
| High school | 3,442 | 3,507 | 3,513 |

#### Test-Taking Rates

Although student participation was voluntary for the CSA, all local educational agencies (LEAs) in California were invited to administer this assessment.

LEAs were given the following guidelines to determine whether a student should take the CSA when either of these conditions applied:

* The student is receiving instruction in Spanish in the state of California.
* The student is seeking a measure that recognizes the student’s Spanish reading, writing mechanics, and listening skills (California Department of Education [CDE], 2022).

Table 8.2 presents the test-taking summary by grade level or the high school grade band in the 2021–22 test administration. Table 8.2 also includes a summary of test-taking data from previous CSA administrations with the exception of the 2019–20 test administration, which produced limited data because of the suspension of testing due to the coronavirus disease 2019 pandemic.

What follows are the findings from this analysis:

* The completion rates for the CSA were about 96 percent of registered students in grades three through five, 87 to 96 percent in grades six through eight, and 85 percent in high school.
* The number of students who took the CSA during the 2021–22 test administration is, on average, 18 percent more than students who took the CSA in 2018–19. For grades three through five, about 22 percent more students took the CSA in 2021–22 than in 2018–19; the same comparison shows 19 percent higher for grades six through eight and 11 percent lower for high school.

Table 8.2 Total Number Tested in the 2021–22 Administration

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **2021–22 Total Registered** | **2021–22 Total Completed** | **2021–22 Percent Completed** | **2018–19 Total Registered** | **2018–19 Total Completed** | **2018–19 Percent Completed** | **Percent Difference of Completion Between 2021–22 and 2018–19** |
| Grade 3 | 11,448 | 11,105 | 97.00 | 10,320 | 9,270 | 89.83 | 19.80 |
| Grade 4 | 10,512 | 10,119 | 96.26 | 9,174 | 8,174 | 89.10 | 23.79 |
| Grade 5 | 8,747 | 8,434 | 96.42 | 7,972 | 6,869 | 86.16 | 22.78 |
| Grade 6 | 6,146 | 5,892 | 95.87 | 6,247 | 4,794 | 76.74 | 22.90 |
| Grade 7 | 4,177 | 3,833 | 91.76 | 5,563 | 3,404 | 61.19 | 12.60 |
| Grade 8 | 3,744 | 3,239 | 86.51 | 5,246 | 2,672 | 50.93 | 21.22 |
| High school | 4,133 | 3,513 | 85.00 | 16,076 | 3,944 | 24.53 | -10.93 |
| **Total:** | **48,907** | **46,135** | **94.33** | **60,598** | **39,127** | **64.57** | **17.91** |

A total of 163 LEAs participated in the CSA. Nearly one half of the total number of LEAs—80 out of 163 LEAs—are from the southern region of California. The mean completion rate is the average of the completion rates from all LEAs in a region.

Table 8.3 presents the test-taking rates of each region in California. Note that this table shows the number of LEAs that actually had students complete the test after students were registered for testing. The mean, minimum, and maximum completion rates are summarized at the LEA level.

Table 8.3 Test-Taking Rates by California Region

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Region** | **No. of LEAs** | **Total Students Registered** | **Total Students Completed** | **Mean Completion Rate** | **Minimum Completion Rate** | **Maximum Completion Rate** |
| North | 32 | 3,892 | 3,687 | 86.35 | 0.00 | 100.00 |
| Central | 51 | 11,747 | 10,748 | 80.73 | 0.00 | 100.00 |
| South | 80 | 33,268 | 31,700 | 75.35 | 0.00 | 100.00 |

### Classical Item Analyses

Classical item analyses are conducted to evaluate the performance of all test items with respect to item difficulty, item-total correlation, and distractor analysis. The associated flagging rules of these statistics are used to identify items that are not performing as expected.

#### Classical Item Difficulty Indices (*p*-value and Average Item Score)

Items scored as one (correct) or zero (incorrect) are referred to as dichotomous items. Items scored from zero to some number of points greater than one are called polytomous items.

For dichotomous items, item difficulty is indicated by its *p*-value, which is the proportion of students who answer the item correctly. The range of *p*-values is from 0.00 to 1.00. Items with high *p*-values are easier items; those with low *p*-values are more difficult. Dichotomous items are flagged for review if their *p*-values are above 0.95 (i.e., too easy). Two-choice dichotomous single-select items, three-choice dichotomous single-select items, and all other dichotomous items are flagged as too difficult if their *p*-values are below 0.50, 0.30, and 0.20, respectively.

The formula for the *p*-value for a dichotomous item is presented in equation 8.1. *Refer to the* [*Alternative Text for Equation 8.1*](#_Alternative_Text_for_37) *for a description of this equation.*

 (8.1)

where,

*Xij* is the score (0 or 1) received for a given dichotomous item *i* for student *j*, and

*Ni* is the total number of students who were presented with item *i*.

For polytomous items, the difficulty is indicated by either the average item score (AIS) or *p*-‍value. The AIS can range from 0.00 to the maximum total possible points for an item. Desired AIS values for polytomous items generally fall within the range of 20 percent to 80 percent of the maximum obtainable item score; items with values outside this range are flagged for review. To facilitate the interpretation, the AIS values for polytomous items are often expressed as the proportion of the maximum possible score, which are equivalent to the *p-*values for dichotomous items.

For polytomous items, the *p-*value is defined as presented in equation 8.2. *Refer to the* [*Alternative Text for Equation 8.2*](#_Alternative_Text_for_38) *for a description of this equation.*

 (8.2)

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*.

#### Item-Total Correlation

An important indicator of item discrimination is the item-total correlation, defined as the correlation between student scores on an individual item and student “total” scores on the test.

The item-total correlation statistic describes the relationship between students’ performance on a specific item and students’ performance on the total assessment. It is calculated as the correlation coefficient between the item score and total score—specifically, the polyserial correlation is used as the index of item-total correlation for both polytomous and dichotomous items. Statistically, it is calculated as the correlation between an observed continuous variable and an unobserved continuous variable hypothesized to underlie the variable with ordered categories (Olsson, Drasgow, & Dorans, 1982). The total scale score or the raw score is used as the criterion score for this analysis.

Theoretically, the polyserial correlation ranges from -1.0 (for a perfect negative relationship) to 1.0 (for a perfect positive relationship) and is estimated as presented in equation 8.3. *Refer to the* [*Alternative Text for Equation 8.3*](#_Alternative_Text_for_36) *for a description of this equation.*

 (8.3)

where,

*β* 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

 is the variance of the criterion (for example, 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 *m-*1 regressions.

Acceptable values for this correlation coefficient are positive and greater than 0.20. A relatively high item-total correlation coefficient value is preferred, as it indicates that higher-ability students tend to perform better on the item than lower-ability students. 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 partial credit, the rubric should be reviewed for whether the rubric for the partial credit score category should be revised.

Items with a low percentage (i.e., less than 3 percent) of students obtaining any score point were flagged for review. Such items may pose problems during IRT calibration. They need to be carefully reviewed and may need to be excluded from the item calibration analyses.

#### Omit Rates

If a student views an item, leaves it unanswered, and then goes on to view and answer another item, the missing response is classified as an “omit.” If the student omits an item—that is, leaves the item unanswered—and does not view additional items, the responses for the successive items are classified as “not seen.”

##### Rates for Dichotomous and Polytomous Items

For both dichotomous and polytomous items, examining the omit rate is useful for identifying potential problems with test features such as testing time and item or test layout. Items with high omit rates are flagged for further investigation by content specialists to ensure that no issues are found with these items. Omit rates for polytomous items tend to be higher than for dichotomous items.

#### Completion Rates

Completion rates are included in table 8.3. The criterion for completion is the student having to log on and answer at least one item.

#### Distractor Analyses

Distractor analyses were conducted on selected-response (SR) items (i.e., items that were not constructed response). The statistics for each item included the proportion of students selecting each distractor (incorrect response), computed for the group of all students in the analysis sample, and were also computed separately for the highest-performing 20 percent of students. Items were flagged for review if more high-performing students chose any distractor rather than the key. Such a result indicated that the item may have multiple correct answers or have the wrong key (i.e., the item was miskeyed).

For SR items, the distractor-total correlation describes the relationship between selecting a distractor for a specific item and performance on the total 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 distractor-total correlations not significantly below zero were flagged for review, as these items may have multiple correct answers, be miskeyed, or have other content issues.

#### Summary of Classical Item Analyses Flagging Criteria

An item was flagged for review if the item analysis yielded any of the following results. One item could have multiple flags if the statistics met the flagging criteria:

* **Difficulty flags** indicated extreme values of the proportion-correct (for dichotomous items) or the proportion of the possible maximum points earned (for polytomous items):
* A-flag: A *p-*value below 0.50 for two-choice dichotomous single-select items, below 0.30 for three-choice dichotomous single-select items, or below 0.20 for all other items.
* H-flag: A *p*-value above 0.95 for dichotomous items or above 0.80 for polytomous items.
* A **discrimination flag** (R-flag) indicated that the item did not discriminate effectively between high- and low-ability students. Items with a polyserial correlation less than 0.20 were flagged.
* An **omit flag** (O-flag) indicated an omission rate above 5 percent for dichotomous multiple-choice, single-select items or above 15 percent for all other items.
* A **distractor flag** (P-flag) was used for an item with any distractors having a correlation with the criterion score that is either positive, zero, or negative but not significantly below zero
* A **miskey flag** (D-flag) was used for multiple-choice items when more of the high-ability examinee group—the top 20 percent of examinees on the total assessment—choose any distractor rather than the response keyed as correct.
* An **underrepresented score point flag** (L-flag) was used for any item that had less than 3 percent of the students at any score level.

ETS’ Psychometric Analysis & Research staff and Assessment and Learning Technology Research & Development staff carefully reviewed each of the flagged items during and at the end of the item analyses. All flagged items were also reviewed by California educators at the data review meeting and then summarized for the CDE with recommendations for subsequent analyses.

#### Classical Item Analyses Results

The summary statistics of the classical item analyses, which include the means and ranges of overall item difficulty and item-total correlation for all operational items, are presented in table 8.4 and table 8.5 for grades three through eight and high school. There is a range of item difficulties, with the *p-*values ranging from 0.04 to 0.82 and the average *p*-values ranging from 0.44 to 0.50. The difficulty distribution of the items shows that the largest group of items is in the range of 0.4 to 0.6, the second largest group of items is in the range of 0.2 to 0.4, and the third largest group of items is in the range of 0.6 to 0.8. Few items are in the highest difficulty group (ranging between 0 and 0.2) and the lowest difficulty group (ranging between 0.8 and 1.0).

[Appendix 8.A](#_Appendix_8.A:_Classical) includes classical item analysis results. Table 8.A.1 presents summary information of item difficulty distribution, the same as table 8.4, broken down by domain and by grade level or the high school grade band. Item information includes *p*-value (item difficulty) intervals, mean and median of *p*-value, minimum and maximum values, as well as the number of items in each domain. Table 8.A.3 through table 8.A.9 present detailed information about individual items for grades three through eight and high school, including *p*-value, item-total-correlation, flag status, maximum score points, and item type.

Table 8.4 Item Difficulty Distributions by Grade Level or Grade Band

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **0 ≤ *p* < 0.2** | **0.2 ≤ *p* < 0.4** | **0.4 ≤ *p* < 0.6** | **0.6 ≤ *p* < 0.8** | **0.8 ≤ *p* ≤ 1.0** | **Total Number of Items** | **Mean *p*-value** | **Minimum *p*-‍value** | **Maximum *p*-‍value** |
| Grade 3 | 0 | 20 | 28 | 4 | 0 | 52 | 0.44 | 0.20 | 0.79 |
| Grade 4 | 1 | 19 | 24 | 8 | 0 | 52 | 0.45 | 0.20 | 0.73 |
| Grade 5 | 2 | 18 | 22 | 9 | 1 | 52 | 0.46 | 0.19 | 0.82 |
| Grade 6 | 0 | 11 | 28 | 12 | 1 | 52 | 0.50 | 0.21 | 0.81 |
| Grade 7 | 1 | 18 | 24 | 9 | 0 | 52 | 0.46 | 0.04 | 0.77 |
| Grade 8 | 0 | 14 | 28 | 9 | 1 | 52 | 0.48 | 0.25 | 0.80 |
| High school | 1 | 17 | 28 | 6 | 0 | 52 | 0.44 | 0.16 | 0.69 |

Overall item-total correlation distributions are presented in table 8.5 and table 8.A.2. Across grade levels, the mean item-total correlations were 0.39 or higher. Only one item from the high school form had negative item-total correlation. This item was revised and will be field-tested again in a future form during a future administration. Summary item-total correlation distributions by domain and by grade level or the high school grade band are presented in table 8.A.2. Detailed item-total correlation information for each item is presented in table 8.A.3 through table 8.A.9 for grades three through eight and high school.

Table 8.5 Item Total Correlation Distributions by Grade Level or Grade Band

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **r < 0** | **0 ≤ r < 0.2** | **0.2 ≤ r < 0.3** | **0.3 ≤ r < 0.4** | **0.4 ≤ r < 0.5** | **r ≥ 0.5** | **Total Number of Items** | **Mean r** | **Minimum r** | **Maximum r** |
| Grade 3 | 0 | 3 | 7 | 12 | 15 | 15 | 52 | 0.43 | 0.13 | 0.73 |
| Grade 4 | 0 | 1 | 10 | 15 | 12 | 14 | 52 | 0.42 | 0.11 | 0.70 |
| Grade 5 | 0 | 1 | 8 | 18 | 13 | 12 | 52 | 0.42 | 0.19 | 0.74 |
| Grade 6 | 0 | 3 | 6 | 10 | 16 | 17 | 52 | 0.44 | 0.11 | 0.70 |
| Grade 7 | 0 | 2 | 11 | 15 | 14 | 10 | 52 | 0.39 | 0.09 | 0.67 |
| Grade 8 | 0 | 2 | 5 | 18 | 13 | 14 | 52 | 0.43 | 0.15 | 0.65 |
| High school | 1 | 0 | 10 | 11 | 20 | 10 | 52 | 0.40 | -0.14 | 0.64 |

The summary of flagged items in each test form by grade level or grade band is presented in table 8.6. Note that there are 52 items on each form. All flagged items were reviewed by content experts and then summarized for, and reviewed by, the CDE. None of the flagged items were identified as having any content flaws during the thorough review by content experts and the CDE.

Table 8.6 Flagged Items Summary in Each Form by Grade Level or Grade Band

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Form** | **No. of Items** | **No. of Flag A Items** | **Percent of Flag A Items** | **No. of Flag H Items** | **Percent of Flag H Items** | **No. of Flag R Items** | **Percent of Flag R Items** | **No. of Flag D Items** | **Percent of Flag D Items** | **No. of Flag P Items** | **Percent of Flag P Items** | **No. of Flag O Items** | **Percent of Flag O Items** | **No. of Flag L Items** | **Percent of Flag L Items** |
| Grade 3 | 1 | 52 | 0 | 0.0 | 0 | 0.0 | 3 | 5.8 | 1 | 1.9 | 2 | 3.9 | 0 | 0.0 | 0 | 0.0 |
| Grade 4 | 1 | 52 | 2 | 3.9 | 0 | 0.0 | 1 | 1.9 | 4 | 7.7 | 3 | 5.8 | 0 | 0.0 | 0 | 0.0 |
| Grade 5 | 1 | 52 | 2 | 3.9 | 0 | 0.0 | 1 | 1.9 | 4 | 7.7 | 3 | 5.8 | 0 | 0.0 | 0 | 0.0 |
| Grade 6 | 1 | 52 | 0 | 0.0 | 1 | 1.9 | 3 | 5.8 | 2 | 3.9 | 3 | 5.8 | 0 | 0.0 | 0 | 0.0 |
| Grade 7 | 1 | 52 | 1 | 1.9 | 0 | 0.0 | 2 | 3.9 | 2 | 3.9 | 7 | 13.5 | 0 | 0.0 | 0 | 0.0 |
| Grade 8 | 1 | 52 | 0 | 0.0 | 0 | 0.0 | 2 | 3.9 | 1 | 1.9 | 9 | 17.3 | 0 | 0.0 | 0 | 0.0 |
| High school | 1 | 52 | 1 | 1.9 | 0 | 0.0 | 1 | 1.9 | 3 | 5.8 | 7 | 13.5 | 0 | 0.0 | 0 | 0.0 |

Detailed results of the classical item analyses for each item by grade level or the high school grade band are presented in [appendix 8.A](#_Appendix_8.A:_Classical). The summary statistics of item difficulty and item-total correlation coefficient by domain in each grade level or grade band are presented in table 8.A.1 and table 8.A.2. The item difficulty and item-total correlation for each item in the forms for each grade level or grade band are listed in table 8.A.3 through table 8.A.9. The maximum score points, item type, and associated item flag information for each item are also presented.

### Differential Item Functioning Analyses

DIF is used to evaluate the consistency of individual item performance for students in different demographic student groups who have the same level of domain performance. For example, DIF evaluates whether female and male students matched to have the same test score perform similarly on each item in the test.

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 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.

There are many possible reasons for DIF. The wording of an item, for example, may be such that one group interprets the question differently than the other, or the reading demands of an item are such that, although reading is not being measured (e.g., in a mathematics test), reading differences between the groups lead to differential outcomes on the item.

DIF analyses were conducted on each test for designated comparison groups. Groups are defined on the basis of demographic variables, such as gender, race or ethnicity, and primary disabilities, if the number of students in the group meets the sample size requirements. These comparison groups are specified in table 8.7.

Table 8.7 Student Groups for DIF Comparison

|  |  |  |
| --- | --- | --- |
| **DIF Type** | **Reference Group** | **Focal Group** |
| Gender | Male | Female |

#### 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 8.4 (Dorans & Holland, 1993). The resulting estimate is interpreted as the relative probability of success on a particular item for members of two groups when matched on ability. *Refer to the* [*Alternative Text for Equation 8.4*](#_Alternative_Text_for_39) *for a description of this equation.*

Equation 8.4; a link to the long description for this equation is found in the preceding paragraph. (8.4)

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 8.5 (Holland & Thayer, 1985). *Refer to the* [*Alternative Text for Equation 8.5*](#_Alternative_Text_for_40) *for a description of this equation.*

 (8.5)

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 8.6.

*Refer to the* [*Alternative Text for Equation 8.6*](#_Alternative_Text_for_35) *for a description of this equation.*

 (8.6)

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 8.8; the flagging criteria for polytomous items are provided in table 8.9. The determination of all significant differences is based on *p*-value < 0.05.

Table 8.8 DIF Categories for Dichotomous Items

|  |  |
| --- | --- |
| **DIF Category** | **Criteria** |
| A (negligible) | * Absolute value of MH D-DIF is less than one or is not significantly different from zero. * Positive values are classified as “A+” and negative values as “A-.” |
| B (moderate) | * Absolute value of MH D-DIF is significantly different from zero but not from one and is at least one; *or* absolute value of MH D-DIF is significantly different from one but is less than 1.5. * Positive values are classified as “B+” and negative values as “B-.” |
| C (large) | * Absolute value of MH D-DIF is at least 1.5 and is significantly different from one. * Positive values are classified as “C+” and negative values as “C-.” |

Table 8.9 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

#### Differential Item Functioning Analysis Results

Summarized DIF results are presented in table 8.10 for grades three through eight and for high school.

No items were identified with C-level DIF across grade levels and the high school grade band. No items were identified with B-level DIF in grades three through eight and only two items showed B-level DIF in high school. A vast majority of items exhibited A-level DIF (no DIF).

Table 8.10 Gender DIF Classifications Summary by Grade Level or Grade Band

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DIF Category** | **Grade 3 Number** | **Grade 3 Percent** | **Grade 4 Number** | **Grade 4 Percent** | **Grade 5 Number** | **Grade 5 Percent** | **Grade 6 Number** | **Grade 6 Percent** | **Grade 7 Number** | **Grade 7 Percent** | **Grade 8 Number** | **Grade 8 Percent** | **High School Number** | **High School Percent** |
| C- | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| B- | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 2 | 3.85 |
| A- | 27 | 51.92 | 26 | 50.00 | 22 | 42.31 | 25 | 48.08 | 26 | 50.00 | 26 | 50.00 | 26 | 50.00 |
| A+ | 25 | 48.08 | 26 | 50.00 | 30 | 57.69 | 27 | 51.92 | 26 | 50.00 | 26 | 50.00 | 24 | 46.15 |
| B+ | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| C+ | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Small N | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| **Total:** | **52** | **100.00** | **52** | **100.00** | **52** | **100.00** | **52** | **100.00** | **52** | **100.00** | **52** | **100.00** | **52** | **100.00** |

### 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 and one or more features of the items, such as its difficulty or discrimination. IRT can be used to calibrate items, link item parameter estimates, scale or equate test scores across different forms or test administrations, evaluate item performance, build an item bank, and assemble test forms.

This section describes how IRT models were used to calibrate and equate operational forms onto the base IRT scale established during the 2018–19 administration as well as to link the field test items onto the base scale. Items that were rejected by both the data review committees and the CDE were typically not included in the calibration process. Exceptions may be made when such items are needed to meet the test blueprint and the item has the correct key but was rejected for other reasons.

#### Item Response Theory Model

The one-parameter logistic item response theory (1PL-IRT) model was used for the item calibration and was selected after consultation with the CDE. In particular, the generalized partial credit model (GPCM) (Muraki, 1992) restricted for 1PL-IRT, which is essentially the partial credit model (PCM) (Masters, 1982), was applied to both dichotomous and polytomous items.

The mathematical form of the GPCM is presented in equation 8.7. *Refer to the* [*Alternative Text for Equation 8.7*](#_Alternative_Text_for_41) *for a description of this equation.*

 (8.7)

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 and is fixed to 0.588 for every item,

*bi* is the location parameter for item *i*,

*div* is the category parameter for item *i* on item 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 8.7 becomes an expression of the one-parameter logistic (1PL) model for dichotomous items.

#### Data Preparation

Prior to IRT calibration analyses, ETS psychometricians reviewed the results of the classical item analyses to decide whether any items were of poor quality and needed to be removed from calibration. The results also were reviewed by ETS content experts and the CDE. The decision whether to remove items from calibration was made in consultation with the CDE.

For IRT calibration, scored item response data was used to create the IRT analysis input data files for each grade level. The IRT analysis input data file was a full matrix containing item-level scores for students who answered at least one item.

Similar to the classical item analyses, “omit” items were treated as incorrect and “not presented” items were treated as blank.

#### Equating

Equating is a procedure where test scores, from different test forms assembled on the basis of the same specifications, are placed onto a reference scale so that scores from different test administrations are comparable. There are two approaches to equate the test forms: preequating and postequating.

A preequating design allows for conversion tables that describe the relationship between raw scores and scale scores, or theta scores and scale scores, to be established prior to the current test administration using data from prior administrations. Preequating relies on having a well-calibrated item bank, robust embedded field-testing processes, and stability in item performance over time.

A postequating design uses the data from the current administration to establish the raw-to-scale-score relationship for the current administration’s form.

Both preequating and postequating involve a common‑item nonequivalent groups design (Kolen & Brennan, 2004).

For all assessments, regardless of whether they are preequated or postequated, IRT calibration and linking were conducted to put the field test item parameters onto the base IRT scale.

The CSA was postequated to the baseline scale established in the 2020–21 administration using the data from the 2021–22 administration.

##### Calibration

After the 2021–22 CSA administration, all operational items within each test (grade level and content area) were calibrated using all available data.

FlexMIRT (Cai, 2017), a multilevel and multiple-group IRT software package for item analysis and test scoring, was used for item calibration analysis. This software can fit a variety of IRT models to both single-level and multilevel data that are dichotomous, polytomous, or both, and was chosen for its superior flexibility among IRT software programs.

The evaluation of the calibration results includes the following steps:

1. Reviewing the item parameter estimates to examine whether these estimates were reasonable
   1. At the form level, the summary statistics for the *b*-parameter estimates (location difficulty) and *d*-parameter estimates (step parameter) were examined, including the mean, SD, median, minimum, maximum, and goodness-of-fit.
   2. At the item level, statistics of individual items were examined, including item difficulty estimates, model-fit statistics, and the IRT-based item parameters.
2. Flagging items that did not perform as expected (All flagged items were discussed thoroughly with the CDE to decide whether those items should be removed from calibration or whether the scoring categories need to be collapsed.)

The calibration process was paralleled by two ETS psychometricians to ensure quality and accuracy of results. Specifically, two psychometricians independently created flexMIRT control files and ran the same input data files and then compared the calibration results. Any differences in the output were investigated. Refer to section [*9.6 Quality Control of Psychometric Processes*](#_Quality_Control_of) for more details of this procedure.

##### Equating of the Operational Form

The new items in the tests for each grade level and the grade band were linked to a calibrated item pool using a common-item nonequivalent groups design (Kolen & Brennan, 2004). The base scales for the CSA were established on the basis of data from the 2020–‍21 administration. The linking of the new operational form onto the 2020–21 base scale is done through a set of linking items (i.e., anchor set) selected from the calibrated item pool and readministered in the current test administration for each grade level or grade band.

After IRT calibration was performed with the 2021–22 test administration’s items, the complete set of anchor items was used to calculate the linking constants to place the item parameters onto the 2020–21 scale by using the mean-to-mean method described in the next subsection. The linking process was carried out iteratively by inspecting differences between the 2021–22 item estimates and base estimates for the anchor items and by removing items for which the item difficulty estimates changed significantly; this is called the robust-z procedure. Robust-z is also described in more detail in subsection [*8.4.3.2.2 Robust-Z Procedure*](#_Robust-Z_Procedure).

###### Mean-to-Mean Transformation

The item difficulty estimates from a new form of calibration may not be comparable to those from the 2020–21 calibration. The difficulty estimates based on a typical year’s data need to be transformed onto the base scale to make them comparable to the item bank parameters.

The mean-to-mean transformation assumes that the item bank and the new form difficulty values differ by a constant; that is, the item bank and the new form difficulty values can be made comparable by adding the same constant for all items. If this assumption is correct, then that constant is the difference between the means of the anchor items from the item bank and the new form difficulty values for the anchor items.

An iterative procedure was implemented to calculate the linking constants using common items in the item bank and the typical year’s administration. For each iteration of linking constants computation, the procedure described in subsection [*8.4.3.2.2 Robust-Z Procedure*](#_Robust-Z_Procedure)is intended to inspect the differences between the transformed new (current administration) and base (2020–21) estimates for the anchor items and remove anchor items for which the item difficulty estimates changed significantly.

There were eight steps involved in making mean-to-mean transformation:

1. Identify the anchor items in both the item bank (2020–21 administration) and the 2021–22 administration.
2. Obtain the item difficulty parameters (*b*-values) of these anchor items that are on the base scale from the item bank.
3. Obtain the item difficulty parameters (*b*-values) of these anchor items from the calibration of the new form.
4. Calculate the average item difficulty for the anchor set on the base scale.
5. Calculate the average item difficulty for the anchor set from the calibration of the new form.
6. Obtain the transformation constant by taking the difference between the two average item difficulties (*b*-values)—using the average item difficulty for the anchor set on the base scale and subtracting the average item difficulty for the anchor set from the calibration of the new form.
7. Obtain a set of adjusted item difficulty parameters (*b*-values) by applying the linking constant to the item difficulty parameters of the anchor items from the new form.
8. Remove anchor items by following the procedure as described in subsection [*8.4.3.2.2 Robust-Z Procedure*](#_Robust-Z_Procedure). The iteration process continues by removing one unstable anchor in each round until no additional items are identified with significant differences between the item difficulty estimates for adjusted new and base items.

###### Robust-Z Procedure

To identify any unstable anchor items, ETS used an outlier detection procedure based on the robust-z statistic (Huynh, 2000; Huynh & Rawls, 2009). In this application, robust-z, as described in equation 8.8, was calculated based on the distribution of the difficulty difference for the anchor items between the item bank and the new form in a typical-year administration. *Refer to the* [*Alternative Text for Equation 8.8*](#_Alternative_Text_for_42) *for a description of this equation.*

 (8.8)

where,

*D* is the difference between the base and transformed new item difficulty of an anchor item;

*MdD* is the median of a distribution of *D* for all anchor items; and

*IQR* is the interquartile range of a distribution of *D* for all anchor items, which is defined as the difference between the third quartile (Q3) and the first quartile (Q1) when all the *D* values are rank ordered.

A large value of this statistic for any anchor item indicates that the reference item difficulty parameter and the new form item difficulty parameter for that item differed substantially.

The criterion for removing anchor items is that the robust-z value is greater than 1.645. One anchor item was removed at each iteration. The following criteria were evaluated at each iteration:

* The correlation between the reference item difficulty estimates and new form item difficulty estimates for the anchor sets should be no less than .95.
* The ratio of standard deviations (RSD) of the reference item difficulty estimates and the new form item difficulty estimates for the anchor items should be between .90 and 1.1.

After each iteration, the mean difference of the anchor sets between the base item-difficulty estimates and the new form item difficulty estimates was recomputed based on the remaining anchor items. Once the final anchor item set was obtained and the linking constant was calculated, ETS will evaluate the percentage of the final anchor items in the form. It is desired that the final anchor set is at least 40 percent of all items in the form. When the equating work was completed, ETS discussed the equating results with the CDE and received approval from the CDE. Removed anchor items were not used in the computation of the linking constants but were still included in calibration and for deriving raw-to-theta conversions.

After equating, the item parameters were linked to the base IRT scale. The raw-to-scale-score conversion table can be established using these parameter estimates. For detailed information on the method to establish the raw-to-scale-score conversion table, refer to subsection [*8.4.6 Scaling the Scores*](#_Scaling_the_Scores_1)*.*

###### Evaluation of Linking

Three indices are used to evaluate the quality of the linking procedure. The first two are the RSD of the two sets of item difficulty estimates for the anchor items (i.e., the item bank estimates and the new form item calibration estimates); and the correlation between the two sets of item difficulty estimates for the anchor items (Huynh, 2009). If the correlation is at least 0.95 and the RSD is between 0.90 and 1.1, the linking results are considered acceptable, and all anchor items are regarded as stable in the linking process. The last index to evaluate the quality of the linking procedure is the percentage of anchor items in the form. The CDE requests a 40 percent overlap between anchor items and all items from the form after linking is complete.

Table 8.11 shows a summary of the procedure of evaluation of anchor items described previously for grades three through eight and high school, which includes the number of all anchor items at the beginning, the number of anchor items removed as a result of mean-to-mean transformation and robust-z procedure, the number of remaining anchor items, and the linking constants of the final iteration of the test for each grade level or grade band.

Table 8.11 Final Linking Summary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Number of Items in Initial Anchor Set** | **Number of Items Removed from the Anchor Set** | **Number of Items in Final Linking Set** | **Linking Constant** |
| Grade 3 | 26 | 4 | 22 | -0.2839 |
| Grade 4 | 31 | 6 | 25 | -0.1700 |

Table 8.11 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Number of Items in Initial Anchor Set** | **Number of Items Removed from the Anchor Set** | **Number of Items in Final Linking Set** | **Linking Constant** |
| Grade 5 | 28 | 3 | 25 | -0.1123 |
| Grade 6 | 26 | 4 | 22 | -0.0292 |
| Grade 7 | 28 | 5 | 23 | -0.0356 |
| Grade 8 | 32 | 4 | 28 | -0.0298 |
| High school | 30 | 8 | 22 | -0.0811 |

Table 8.12 presents the total number of operational items on the general form for grades three through eight and high school, the number of remaining anchor items after robust-z evaluation, the percentage of remaining anchor items out of all the operational items on the general form, the correlation between the final anchor set of the 2021–22 item estimates and the reference (2018–19 test administration) difficulty estimates for the anchor items, and the RSD between the final set of the 2021–22 item estimates and the reference (2018–19 test administration) difficulty estimates for the anchor items.

As table 8.12 shows, the correlation is 0.96 or higher and the RSD is between 0.96 and 1.08. The percentage of anchor items in a form is 42 or higher across grade levels and the high school grade band. All three indices to evaluate the quality of equating are met for grades three through eight and high school.

Table 8.12 Evaluation of Anchor Set (Common Items) Between 2018–19 and 2021–22

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Number of Operational Items** | **Anchor Items Remaining After Deletions** | **Remaining Anchor Items as a Percentage of All Operational Items** | **Correlation Between Item Bank Item Parameters and 2021–‍22 Item Parameters** | **RSD of Item Parameters Between Item Bank and 2021–‍22 Administration** |
| Grade 3 | 52 | 22 | 42% | 0.96 | 1.08 |
| Grade 4 | 52 | 25 | 48% | 0.99 | 1.03 |
| Grade 5 | 52 | 25 | 48% | 0.99 | 1.01 |
| Grade 6 | 52 | 22 | 42% | 0.98 | 1.03 |
| Grade 7 | 52 | 23 | 44% | 0.97 | 1.06 |
| Grade 8 | 52 | 28 | 54% | 0.98 | 0.99 |
| High school | 52 | 22 | 42% | 0.99 | 0.96 |

Table 8.13 presents the summary statistics of the final linking results after items with unstable parameters were detected and removed from the anchor set in grades three through eight and high school. The statistics include the number of remaining items in the final anchor set, average item difficulties of the anchor set both in the 2018–19 base scale and from the 2021–22 test administration, along with their differences, as well as the criteria for evaluating the absolute differences.

Table 8.13 Linked Item Parameter Results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Number of Anchor Items** | **Operational Baseline Scale Average *b*-‍parameter** | **Linked 2021–‍22 Test Average *b*-‍parameter** | **Difference of Average *b*-‍parameters** | **Criteria for the Acceptable Absolute Difference** |
| Grade 3 | 26 | 0.042 | 0.042 | 0.0000 | < 0.1 |
| Grade 4 | 31 | 0.087 | 0.087 | 0.0000 | < 0.1 |
| Grade 5 | 28 | 0.070 | 0.070 | 0.0000 | < 0.1 |
| Grade 6 | 26 | -0.027 | -0.027 | 0.0000 | < 0.1 |
| Grade 7 | 28 | 0.150 | 0.150 | 0.0000 | < 0.1 |
| Grade 8 | 32 | -0.197 | -0.197 | -0.0001 | < 0.1 |
| High school | 30 | -0.178 | -0.178 | 0.0001 | < 0.1 |

#### Calibration and Linking for the Field Test Items

The CSA uses the embedded field test design to update item pools for future form assembly. Thirty embedded field test items were tested per grade level in grades three through eight, and 48 embedded field test items were tested in the high school form. Among the 30 field test items, each 10-item set for grades three through eight, or each 16-item set for high school, was embedded in one of the three versions of the test form per grade level or the high school grade band. Under the same design, 46 embedded field test items were included for high school, because one field test item was across the three versions.

The procedures for field test calibration and scaling were as follows:

1. Generate a sparse matrix with all operational items and field test items for each grade level or grade band, including all item scores
2. Calibrate the embedded field test items and the operational items using flexMIRT software
3. Scale the calibrated field test items to the base scale by updating the field test item estimates using the calculated linking constant based on all operational items (In this process, the robust-z and mean-to-mean procedures mentioned previously were applied.)
4. Update the item pools

##### Calibration

The calibration will be conducted using a sparse matrix combining all operational items and field test items from all versions within a grade level. Refer to subsection [*8.4.1 Item Response Theory Model*](#_Item_Response_Theory_1) for the IRT models and the software used in the calibration, subsection [*8.4.2 Data Preparation*](#_Data_Preparation) for the creation of the sparse matrix.

FlexMIRT (Cai, 2017), a multilevel and multiple-group IRT software package for item analysis and test scoring, was used for item calibration analysis. This software can fit a variety of IRT models to both single-level and multilevel data that are dichotomous, polytomous, or both, and was chosen for its superior flexibility among IRT software programs.

The evaluation of the calibration results includes the following steps:

1. Reviewing the item parameter estimates to examine whether these estimates were reasonable
   1. At the form level, the summary statistics for the *b*-parameter estimates (location difficulty) and *d*-parameter estimates (step difficulty) were examined, including the mean, SD, median, minimum, maximum, and goodness-of-fit.
   2. At the item level, statistics of individual items were examined, including item difficulty estimates, model-fit statistics, and the IRT-based item parameters.
2. Flagging items that did not perform as expected (All flagged items were discussed thoroughly with the CDE to decide whether those items should be removed from calibration or whether the scoring categories need to be collapsed.)

The calibration process was paralleled by two ETS psychometricians to ensure quality and accuracy of results. Specifically, two psychometricians independently created flexMIRT control files and ran the same input data files and then compared the calibration results. Any differences in the output were investigated. Refer to section [*9.6 Quality Control of Psychometric Processes*](#_Quality_Control_of)for more details of this procedure.

##### Linking

The Stocking-Lord (1983) method will be applied for the linking. The software STUIRT (Kim & Kolen, 2004) is employed to find the Stocking-Lord linking constants for the common items. The stability of the anchor items will be evaluated by means of item response functions (IRFs) and a d-square measure. The d-square measure is the weighted squared deviation of the IRFs across the range of proficiency (i.e., theta or ) and under a hypothetical normal distribution for between the item parameter estimates from the calibration and the equated values. For a given item , the measure  or “d-squared” is defined in equation 8.9. *Refer to the* [*Alternative Text for Equation 8.9*](#_Alternative_Text_for_43) *for a description of this equation.*

 (8.9)

where,

*i* indexes common or anchor items,

*k* indexes quadrature points for *θ*,

*Pi,new* is the probability of a correct response to item *i* under the calibration of operational items from the current administration,

While *Pi,bank* is the probability of a correct response to item *i* based on the item bank parameter estimates, and

*g(θk)* are weights for each of the *k* quadrature points *θk*.

There are 41 quadrature points from -4.00 to 4.00, and the weights are normally distributed.

The magnitude of d-square in conjunction with plots of these curves are used to identify items for removal from the anchor set. The frequently used minimum d-squared criteria for anchor item removal are 0.1252 for dichotomous items and 0.22 for polytomous items. For anchor items with d-squared results that are at the borderline of these thresholds, the evaluation will be made on an individual basis.

The item parameters obtained through the calibration are on a different scale and will be linked to the baseline scale using all operational items as anchors. The mean-to-mean linking procedures was used and the robust-z method was used to check the stability of the anchors. Refer to subsection [*8.4.3.2.1 Mean-to-Mean Transformation*](#_Mean-to-Mean_Transformation) and subsection [*8.4.3.2.2 Robust-Z Procedure*](#_Robust-Z_Procedure) for more details on these methods.

#### Parameter Estimates

The overall summary of the linked IRT *b*-value estimates for the 2021–22 CSA administration calibration is shown in table 8.14. The mean, SD, minimum, and maximum values are presented, in addition to the number of items for each grade level and the high school grade band. The root mean square error of approximation (RMSEA) values are also provided in table 8.14, which are 0.03, below 0.05 for all grade levels and the grade band, indicating good model fit (Browne & Cudeck, 1993).

Table 8.14 IRT Summary *b*-value Estimates for All CSA Operational Items

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Number of Items** | **Average of *b-*value** | **SD *b*-‍value** | **Minimum *b*-value** | **Maximum *b*-value** | **RMSEA** |
| Grade 3 | 52 | 0.25 | 0.58 | -1.61 | 1.51 | 0.03 |
| Grade 4 | 52 | 0.19 | 0.66 | -1.45 | 1.51 | 0.03 |
| Grade 5 | 52 | 0.18 | 0.71 | -1.65 | 1.52 | 0.03 |
| Grade 6 | 52 | -0.03 | 0.68 | -1.50 | 1.44 | 0.03 |
| Grade 7 | 52 | 0.19 | 0.73 | -1.29 | 3.27 | 0.03 |
| Grade 8 | 52 | 0.04 | 0.62 | -1.53 | 1.19 | 0.03 |
| High school | 52 | 0.29 | 0.62 | -1.24 | 1.75 | 0.03 |

All *b*-values were between -2.0 and +3.5. The average *b-*parameters for the CSA in grades three through five and grade seven through high school were above zero, indicating that, in general, the items were relatively difficult for the students in those grade levels. The mean *b*-parameters for grade six were below zero, indicating the items, on average, were relatively easier for students in that grade level.

The *b*-parameter estimates and *d*-parameter estimates for each item are shown in [appendix 8.B](#_Appendix_8.B:_Item), in table 8.B.1 through table 8.B.7, for grades three through eight and for high school.

Table 8.15 presents the item difficulty parameter distribution by grade level or the high school grade band. The number of items in each of the *b*-parameter intervals is shown for grades three through eight and for high school.

Table 88.15 Item Difficulty Parameter Distribution by Grade Level or Grade Band

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **IRT-b Range** | **Grade 3** | **Grade 4** | **Grade 5** | **Grade 6** | **Grade 7** | **Grade 8** | **High School** |
| b < −3.5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| −3.5 ≤ b < −3.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| −3.0 ≤ b < −2.5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| −2.5 ≤ b < −2.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| −2.0 ≤ b < −1.5 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| −1.5 ≤ b < −1.0 | 1 | 5 | 3 | 5 | 2 | 2 | 1 |
| −1.0 ≤ b < −0.5 | 7 | 4 | 8 | 8 | 6 | 8 | 7 |
| −0.5 ≤ b < 0 | 22 | 13 | 7 | 11 | 14 | 12 | 11 |
| 0 ≤ b < 0.5 | 11 | 17 | 18 | 18 | 13 | 18 | 16 |
| 0.5 ≤ b < 1.0 | 8 | 9 | 11 | 6 | 13 | 9 | 10 |
| 1.0 ≤ b < 1.5 | 2 | 3 | 4 | 3 | 3 | 2 | 6 |
| 1.5 ≤ b < 2.0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 2.0 ≤ b < 2.5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2.5 ≤ b < 3.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3.0 ≤ b < 3.5 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| b ≥ 3.5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Min | -1.91 | -1.63 | -1.78 | -1.54 | -1.33 | -1.56 | -1.33 |
| Max | 1.23 | 1.35 | 1.43 | 1.43 | 3.24 | 1.17 | 1.68 |
| Mean | -0.03 | 0.02 | 0.07 | -0.06 | 0.16 | 0.01 | 0.21 |
| SD | 0.59 | 0.67 | 0.72 | 0.68 | 0.73 | 0.62 | 0.63 |
| **Number of Items:** | **52** | **52** | **52** | **52** | **52** | **52** | **52** |

Table 8.B.1 through table 8.B.7 in [appendix 8.B](#_Appendix_8.B:_Item) provide the linked IRT difficulty and step parameter estimates at the item level in each grade level or the high school grade band.

#### Scaling the Scores

The raw scores on each new form were transformed to scale scores on the reference scale using a two-step procedure. First, the number-correct scores (raw scores) were transformed to ability (theta) scores on the reference scale by the inverse test characteristic curve (TCC) procedure described in the next subsection. Then, these ability (theta) scores were transformed to scale scores through the linear transformation described in subsection [*8.4.6.2 Transformation from Theta Scores to Scale Scores*](#_Transformation_from_Theta_1).

##### Inverse Test Characteristic Curve Procedure

After all the item difficulty estimates are transformed to the reference scale, students’ overall ability estimates can be derived from the input data file that was described in subsection [*8.4.2 Data Preparation*](#_Data_Preparation)*,* through the IRT inverse TCC method (Stocking, 1996). This method transforms the sum of the student’s item scores into an ability estimate. That estimate is the ability value that makes the sum of the expected scores on the items administered to the student equal to the sum of the scores that the student actually received on those items.

The TCC expresses the expected total score on a set of items as a function of the student’s ability, which is shown in equation 8.10. *Refer to the* [*Alternative Text for Equation 8.10*](#_Alternative_Text_for_44) *for a description of this equation.*

 (8.10)

where,

*i* indexes dichotomous items,

*j* indexes polytomous items,

*ndich* is the number of dichotomous items in the test,

*Pi(θ)* is the probability of a correct response to item *i* at ability *θ* on the dichotomous item in equation 8.7,

*npoly* is the number of polytomous items in the test,

*m* is the number of score categories for each polytomous item,

*sxj* is the value for score category x for the polytomous item *j*,

*Pxj(θ)* is the probability that an examinee with ability *θ* obtains score sx on the polytomous item *j* in [equation 8.7](#EQ87), and

*ξ(θ)* is the corresponding expected total score.

##### Transformation from Theta Scores to Scale Scores

Students’ ability estimates (theta scores) were transformed to the scale score metric by applying a linear transformation based on threshold theta values. Those threshold values were determined after standard setting and approved by the California State Board of Education (SBE). The scaling transformation from theta to scale score is shown in equation 8.11. *Refer to the* [*Alternative Text for Equation 8*.*11*](#_Alternative_Text_for_45) *for a description of this equation.*

 (8.11)

where,

*SS* represents the reporting scale score,

*θ* represents student ability,

*a* represents the slope of the linear transformation, and

*b* represents the intercept of the linear transformation.

The slope is calculated using equation 8.12. *Refer to the* [*Alternative Text for Equation 8.12*](#_Alternative_Text_for_19) *for a description of this equation.*

 (8.12)

The intercept is calculated in equation 8.13 with range 3 thresholds. *Refer to the* [*Alternative Text for Equation 8.13*](#_Alternative_Text_for_20) *for a description of this equation.*

 (8.13)

where,

*SSthreshold3* represents the reporting scale score range 3 threshold,

represents the SD of the reporting scale score,

is the SD of the theta scores, and

*θthreshold3* is the reporting range 3 threshold theta score.

The slope and intercepts derived from equations 8.12 and 8.13 defined the linear relationship between the theta scale and the scale-score scale based on the 2018–19 data. *Refer to the* [*Alternative Text for Equation 8.14*](#_Alternative_Text_for_22) *for a description of this equation.*

 (8.14)

The slopes and intercepts for each grade level or the high school grade band are shown in table 8.16.

Table 8.16 Conversion of Theta Score to Reporting Scores by Grade Level or Grade Band Based on the 2018–19 Test Administration (the Baseline)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade Level or Grade Band | *σθ* | *σSS* | Score Reporting Range 2 Threshold Theta Score | Score Reporting Range 3 Threshold Theta Score | Score Reporting Range 2 Threshold | Score Reporting Range 3 Threshold | Slope *a* | Intercept *b* |
| Grade 3 | 0.732 | 10 | 0.037 | 0.860 | 349 | 360 | 14 | 348 |
| Grade 4 | 0.798 | 10 | 0.045 | 0.922 | 449 | 460 | 13 | 448 |
| Grade 5 | 0.727 | 10 | -0.114 | 0.886 | 546 | 560 | 14 | 548 |
| Grade 6 | 0.692 | 10 | -0.128 | 0.703 | 648 | 660 | 14 | 650 |
| Grade 7 | 0.656 | 10 | 0.128 | 1.179 | 744 | 760 | 15 | 742 |
| Grade 8 | 0.681 | 10 | 0.137 | 0.981 | 848 | 860 | 15 | 846 |
| High school | 0.727 | 10 | 0.171 | 0.916 | 950 | 960 | 14 | 947 |

### Testing Time Analyses

The length of time it takes students to complete an assessment is recorded and analyzed to build a profile describing what a typical testing event looks like for each grade-level assessment. In addition, variability in testing time is investigated to determine whether a student’s testing time should be viewed as unusual or irregular for further investigation. It should be noted that the CSA is untimed.

In these analyses, all students who completed testing with a valid reporting scale score are included. The testing population is partitioned into performance quartiles based on all operational items. The descriptive statistics—for example, the number of students, mean, SD, minimum and maximum, percentiles—of the time required to complete the total test are computed for each of the four performance quartile groups for each grade level (i.e., grades three through eight and the high school grade levels).

[Appendix 8.C](#_Appendix_8.C:_Testing) summarizes the results of the testing time analysis. Table 8.C.1 through table 8.C.7 provide descriptive statistics of total testing time for the full student population at each ability level for each grade level or the high school grade band for the general and accommodated forms. The four quartiles, Q1 through Q4, are represented by four raw score intervals. The unit of testing time is in minutes; for example, in table 7.C.1, the median (i.e., fiftieth percentile) of the testing time is 60.68 minutes for the grade three Q1 (raw score interval 0–‍22) group of students who took the general form.

In general, students who took the general forms or the accommodated forms at the lowest quartile level (Q1), which is the lowest raw score interval, have shorter testing times than students in the other quartile groups except for the accommodated form for high school, where students in Q1 spent more time than those in Q2. The median total testing time generally increases as the quartile level increases from the first quartile to the last quartile (Q4), the highest raw score interval, meaning that the students who performed better on the CSA tended to spend more time on the test.

### Reliability Analyses

The reliability for a particular group of students’ test scores estimates the extent to which the scores would remain consistent if those same students were retested with a parallel version of the same test. There are many definitions of reliability (Haertel, 2006) that have their genesis in classical test theory and a variety of methods that can be used to estimate reliability.

The general concept of reliability concerns the extent to which the test scores measure *a particular construct* consistently. The variance in the distribution of test scores—essentially, the observed differences among individuals—is partly due to differences that are consistent and partly due to differences that are not consistent. The measure of variation associated with the first kind of differences—consistent differences—is called “true variance”; this would include actual differences in students’ knowledge. The measure of variation associated with the remaining differences—those that operate essentially at random—is called “error variance.” Error variance includes a variety of underlying differences such as selections of test content, which may cause a student’s test score to be slightly higher in one evaluation and slightly lower in another. Reliability is the proportion of total variance that is due to true variance. The standard error of measurement (SEM) is a statistic that characterizes the error variance.

Reliability coefficients range from zero to one. The higher the reliability coefficient for a set of scores, the more likely individuals are to obtain very similar scores upon repeated testing occasions, if the students do not change in their level of the knowledge or skills measured by the test.

#### Sample for Reliability Analyses

The reliability analyses performed for the California Assessment of Student Performance and Progress (CAASPP) require that the sample be screened beyond the requirements listed in subsection [*8.1.2 Samples Used for Analyses*](#_Samples_Used_for).

#### Reliability Measures

The reliability coefficient cannot, in fact, be computed directly unless the student actually takes two parallel versions of the same test. However, with some reasonable assumptions, reliability can be estimated from the students’ responses to a single version of the test.

Like other statistics, the reliability coefficient can vary substantially from one group of students to another. It tends to be larger in groups that are more diverse in the ability measured by the test and smaller in groups that are more homogeneous in the ability measured.

The CAASPP test reliabilities were evaluated for each domain, the composite scores, or for item types as dichotomous items and polytomous items, using the coefficient alpha (Cronbach, 1951) index of internal consistency, which is calculated as presented in equation 8.15. *Refer to the* [*Alternative Text for Equation 8.15*](#_Alternative_Text_for_46) *for a description of this equation.*

 (8.15)

where,

*k* is the number of items on test form,

 is the estimated variance of item *i*, and

 is the estimated total test variance.

The reliability of the overall score was estimated by substituting sample estimates into the following definitional formula for composite reliability (Feldt & Brennan, 1989), equation 8.16. *Refer to the* [*Alternative Text for Equation 8.16*](#_Alternative_Text_for_48) *for a description of this equation.*

 (8.16)

where,

 is the weight of the *j*th component in forming the composite score,

 is the variance of scores on the *j*th component,

 is the reliability of scores on the *j*th component, and

 is the variance of the composite score.

#### Standard Error of Measurement

The SEM is a measure of how much students’ scores would vary from the scores they would earn on a perfectly reliable test. If it were possible to compute the error of measurement for each student’s score in a large group of students, these errors of measurement would have a mean of zero. These SEMs would be an indication of how much the errors of measurement affect the students’ scores. The SEM is expressed in the same units as the test score, whether the units are in raw score or scale score metric. In a large group of students, approximately two-thirds of the students will earn scores within one SEM of the scores they would earn on a perfectly reliable test.

The formula for the SEM is presented in equation 8.17. *Refer to the* [*Alternative Text for Equation 8.17*](#_Alternative_Text_for_47) *for a description of this equation.*

 (8.17)

where,

 is the reliability estimated in equation 8.16, and

 is the SD of the total score.

Table 8.17 gives the reliability and SEMs for each CSA form, along with the number of items and students upon which those analyses were performed. In the table, form 1 represents the general form while form A represents the accommodated form.

Table 8.17 Test Reliability of the Total Scores

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Form** | **No. of Items** | **N Points** | **N Students** | **Mean** | **SD** | **Reliability** | **SEM** |
| Grade 3 | 1 | 52 | 66 | 9,767 | 30.13 | 9.96 | 0.86 | 3.66 |
| Grade 3 | A | 52 | 60 | 1,271 | 22.00 | 7.23 | 0.77 | 3.43 |
| Grade 4 | 1 | 52 | 66 | 9,136 | 31.17 | 9.32 | 0.85 | 3.59 |
| Grade 4 | A | 52 | 61 | 955 | 26.26 | 8.56 | 0.84 | 3.43 |
| Grade 5 | 1 | 52 | 64 | 7,735 | 29.80 | 9.06 | 0.85 | 3.56 |
| Grade 5 | A | 52 | 63 | 670 | 26.66 | 7.40 | 0.78 | 3.47 |
| Grade 6 | 1 | 52 | 65 | 5,380 | 33.67 | 9.78 | 0.86 | 3.63 |
| Grade 6 | A | 52 | 64 | 476 | 31.96 | 8.76 | 0.84 | 3.55 |
| Grade 7 | 1 | 52 | 65 | 3,693 | 30.95 | 8.70 | 0.82 | 3.64 |
| Grade 7 | A | 52 | 63 | 107 | 29.04 | 7.39 | 0.77 | 3.52 |
| Grade 8 | 1 | 52 | 64 | 3,125 | 32.03 | 9.31 | 0.85 | 3.58 |
| Grade 8 | A | 52 | 58 | 82 | 25.39 | 7.63 | 0.80 | 3.39 |
| High school | 1 | 52 | 62 | 3,449 | 28.32 | 8.82 | 0.84 | 3.55 |
| High school | A | 52 | 62 | 13 | 33.92 | 8.18 | 0.85 | 3.21 |

#### Student Group Reliabilities and Standard Errors of Measurement

For the 2021‒22 test administration, CSA reliabilities are estimated for various student groups that tested. The student groups included in these analyses are defined by their gender, economic status, provision of special services, length of attendance in US schools, whether they received instruction in Spanish, and English language acquisition status. Reliabilities and SEM information for the total test scores by test form are reported for each student group analysis.

Reliability values are estimates that approach the true reliability as the number of students whose scores contribute to the estimates increases. Reliabilities are not reported for samples that comprise 10 or fewer students. Results based on samples that contain 50 or fewer students should be interpreted with caution, because these estimates may meaningfully deviate from the true reliability. In some cases, score reliabilities were not estimable and are presented in the tables as “N/A.”

Table 8.D.1 through table 8.D.6 in [appendix 8.D](#_Appendix_8.D:_Reliability_1) present the overall test reliabilities for the various student groups. Most student groups have reliability greater than 0.80 for the general forms across all grade levels and the high school grade band, with the exception of

* students who received special education services,
* English learner (EL) students in grades four through eight, and
* students who were not receiving Spanish instruction for grades three and seven.

Among those groups, the lowest reliability value, 0.70, is for students who received special education services in high school. It should be noted that in this case, the low reliability was likely due to the lack of variation in student performance because of the small number of students in this student group.

#### Conditional Standard Errors of Measurement

Classical test theory assumes that the standard error of a test score is constant throughout the score range. While the assumption is probably reasonable in the mid-score ranges, it is less reasonable at the extremes of the score distribution. IRT expands the concept by providing estimates of the standard error at each score point on the distribution.

##### Methodology

Conditional standard errors of measurement (CSEMs) are estimated as part of the IRT-based scoring procedure. CSEMs for scale scores are based on IRT and are estimated as a function of measured ability. The CSEMs of theta scores (or of linearly transformed theta scores) are smaller at points of the scale in the test metric where more items are located. A student’s CSEM under the IRT framework is equal to the reciprocal of the square root of the test information function (TIF) based on the items taken by each student. The CSEM for a student with proficiency 𝜃𝑗 is calculated using equation 8.18. *Refer to the* [*Alternative Text for Equation 8.18*](#_Alternative_Text_for_49) *for a description of this equation.*

 (8.18)

where,

*I(θj)* is the test information for student *j* and is calculated using equation 8.19. *Refer to the* [*Alternative Text for Equation 8.19*](#_Alternative_Text_for_50) *for a description of this equation.*

 (8.19)

where,

*Ii(θj)* is the item information of item *i* for student *j*.

Item information is calculated as presented in equation 8.20. *Refer to the* [*Alternative Text for Equation 8.20*](#_Alternative_Text_for_51) *for a description of this equation.*

 (8.20)

where,

*si(θj)* is the expected item score for item *i* on a theta score *θj* calculated as presented in equation 8.21 *(Refer to the* [*Alternative Text for Equation 8.21*](#_Alternative_Text_for_52) *for a description of this equation.*)

 (8.21)

and equation 8.22. *Refer to the* [*Alternative Text for Equation 8.22*](#_Alternative_Text_for_27) *for a description of this equation.*

 (8.22)

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 8.7; and

*ni* is the maximum number of score points for item *i*.

CSEMs for scale scores are computed by transforming CSEMs of theta scores onto the reporting scale. Refer to subsection [*8.4.6.2 Transformation from Theta Scores to Scale Scores*](#_Transformation_from_Theta_1) for scaling factors of transformation. A student’s CSEM for scale scores under the IRT framework is equal to the reciprocal of the square root of the TIF multiplied by thescaling factor a, as presented in equation 8.23*. Refer to the* [*Alternative Text for Equation 8.23*](#_Alternative_Text_for_28) *for a description of this equation.*

 (8.23)

where,

;

CSEM(*SS*) is the conditional SEM on scale score scale;

 is the TIF at ability level  as shown in equations 8.20, 8.21, and 8.22; and

*a* is the scaling factor (the slope) needed to transform theta to the scale score metric.

##### Results

Table 8.18 presents the scale score CSEMs at the lowest score required for a student to be classified in the score reporting range 2 and score reporting range 3 for each CSA. CSEMs for each individual score point are presented in table 7.A.1 through table 7.A.7 in [appendix 7.A](#_Appendix_7.A:_Raw) for grades three through eight and high school.

Table 8.18 Scale Score CSEMs at the Score Reporting Range Thresholds

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Level 2―Score Reporting Range Threshold** | **Level 2 CSEM** | **Level 3―Score Reporting Range Threshold** | **Level 3 CSEM** |
| Grade 3 | 349 | 4 | 360 | 4 |
| Grade 4 | 449 | 3 | 460 | 4 |
| Grade 5 | 546 | 4 | 561 | 4 |
| Grade 6 | 648 | 4 | 660 | 4 |
| Grade 7 | 744 | 4 | 760 | 4 |
| Grade 8 | 848 | 4 | 860 | 4 |
| High school | 950 | 4 | 960 | 4 |

#### Decision Classification Analyses

When an assessment uses reporting ranges as the primary method to report test results, accuracy and consistency of decisions become key indicators of the quality of the assessment.

##### Methodology

The reliabilities of reporting range classifications, which are criterion referenced, are related to the reliabilities of the test scores on which they are based; however, they are not exactly the same. Glaser (1963) was among the first to draw attention to this distinction, and Feldt and Brennan (1989) reviewed the topic extensively. While test reliability evaluates the consistency of test scores, decision classification reliability evaluates the consistency of classification.

Decision accuracy is the extent to which students are classified in the same way as they would be if each student’s score were the average over all possible forms of the test (the student’s true score). Decision accuracy answers the following question: How closely does the actual classification of test takers, based on their single-form scores, agree with the classification that would be made on the basis of their true scores, if their true scores could somehow be known?

Decision consistency is the extent to which students are classified in the same way as they would be on the basis of a single form of an assessment other than the one for which data is available. Decision consistency answers the following question: What is the agreement between the classifications based on two nonoverlapping, equally difficult forms of the test?

The methodology used for estimating the reliability of classification decisions is described in Livingston and Lewis (1995). The necessary input information includes only the maximum and minimum possible scores on the assessment and the observed score distribution and the reliability coefficient for the group of students referenced by the estimates. The method was implemented by the ETS proprietary computer program RELCLASS-COMP (Version 4.14).

Reliability of classification at a threshold is estimated by combining the reporting ranges above a particular threshold and combining the reporting ranges below that threshold. The result is a two-by-two table indicating whether the students are above or below the threshold. The sum of the entries in the main diagonal is the number of students accurately (or consistently) classified as above or below that threshold.

Table 8.19 and table 8.20 illustrate these two-by-two contingency tables. The proportion of students being accurately classified is determined by summing across the diagonals of the upper tables. The proportion of consistently classified students is determined by summing the diagonals of the lower tables.

Table 8.19 Decision Accuracy for Reaching a Reporting Range

|  |  |  |
| --- | --- | --- |
| **Reporting Range Status** | **Does Not Reach a Reporting Range Based on True Score** | **Reaches a Reporting Range Based on True Score** |
| Does not reach a reporting range | Correct classification | Incorrect classification |
| Reaches a reporting range | Incorrect classification | Correct classification |

Table 8.20 Decision Consistency for Reaching a Reporting Range

|  |  |  |
| --- | --- | --- |
| **Reporting Range Status** | **Does Not Reach a Reporting Range Based on an Alternate Form** | **Reaches a Reporting Range Based on an Alternate Form** |
| Does not reach a reporting range | Consistent classification | Inconsistent classification |
| Reaches a reporting range | Inconsistent classification | Consistent classification |

##### Results

The results of these analyses are presented in table 8.E.1 through table 8.E.14 in [appendix 8.E](#_Appendix_8.E:_Analyses). Each table pair includes the contingency tables for both accuracy and consistency of the various reporting range classifications. The proportion of students being accurately classified is determined by summing across the diagonals of the first table of the table pairs for decision accuracy (i.e., table 8.E.1, table 8.E.3, etc.). The proportion of consistently classified students is determined by summing the diagonals of the second table of the table pairs for decision consistency (i.e., table 8.E.2, table 8.E.4, etc.).

The overall decision accuracy is greater than 0.80 for all tests—grades three through eight and the high school grade band—with the highest accuracy of 0.88 occurring for grade three and the lowest level of accuracy of 0.81 occurring in grade six. The overall decision consistency is relatively lower, with the lowest consistency of 0.74 occurring for grade six and the highest consistency of 0.83 occurring for grade three.

### 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, standard setting, test scaling and linking, scoring, reporting, and score usage.

In this section, the evidence gathered is presented to support the intended uses and interpretations of scores for the CSA. This section discusses some of 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 relations to other variables
3. Evidence based on response processes
4. Evidence based on internal structure
5. Evidence based on the consequences of testing

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

#### Evidence Based on the Design of the CSA

##### Purposes

The CSA is designed to measure a student’s Spanish skills in reading, writing mechanics, and listening for the purposes of

* providing student-level data in Spanish literacy,
* providing aggregated data that may be used for evaluating the implementation of Spanish reading/language arts programs at the local level, and
* providing a high school measure suitable to be used, in part, for the California State Seal of Biliteracy.

The assessment provides students an annual opportunity to measure their reading/‌language arts literacy in Spanish.

##### The Constructs to Be Measured

As a voluntary assessment to measure a student’s Spanish skills in reading, writing mechanics, and listening, the CSA is designed to show how well students perform relative to the Spanish version of Common Core English language arts/literacy standards (i.e., California Common Core State Standards en Español [CCCSSeE]), which was developed as a joint effort between the San Diego County Office of Education, Council of Chief State School Officers, and the CDE (San Diego County Office of Education, n.d.).

The CCCSSeE are organized into the following domains:

* Reading standards
* Writing standards
* Speaking/Listening standards
* Language standards

It should also be noted that while the focus of the CCCSSeE is acquired language arts literacy, the domains in the previous list are also harmonious with a four-skill, language-learning framework (e.g., listeningand *reading*, known as “receptive” skills, and speakingandwriting,known as “productive” skills).[[5]](#footnote-6)

Test blueprints are used to measure students’ mastery of the standards included in theCCCSSeE (CDE, 2017). They also provide an operational definition of the construct to which each set of standards refers and define the following:

* Subject to be assessed
* Tasks to be presented
* Administration instructions to be given
* 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).

ETS developed all CSA items to conform to the SBE-approved test blueprints.

##### The Interpretations and Uses of the Scores

Overall student performance expressed as scale scores are generated for the CSA. The scale score is also used to classify students in terms of their score reporting range by grade level or the high school grade band.

The grade-level- or grade-band-specific score reporting range descriptors help to communicate the different skills in Spanish reading/language arts knowledge that correspond with one of the three score reporting ranges. The importance of these range descriptors is that they define the knowledge or skill expectations at each range on a functional basis, based on the standards and general achievement level descriptors, and give standardized meaning to scores or score reporting ranges.

A description of the uses and applications of the CSA results is presented in [*Chapter 7*: *Scoring and Reporting*](#_Scoring_and_Reporting).

The CSA 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 schools and 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

The intended test population for the CSA consists of students enrolled in grades three through twelve receiving instruction in Spanish in California and students seeking a measure that recognizes their Spanish-specific academic reading, writing mechanics, and listening skills. It is critical to recognize the diverse characteristics of the test population for the CSA and the contexts in which the test purpose and use are situated.

#### Evidence Based on 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 the State Standards

The CSA is aligned with the CCCSSeE. The purpose of the CCCSSeE is to guide instruction in a multitude of contexts, including in-class, collaborative activities. The focuses of the CCCSSeE are acquired language arts literacy and the necessary knowledge and skills needed to reach the standards in each grade level (San Diego County Office of Education, n.d.).

##### Item Specifications

Item specifications describe the characteristics of items that are written to measure each content standard. ETS maintains item specifications for each grade-level CSA. The specifications for the CSA are described in [*Chapter 3: Item Development and Review*](#_Item_Development_and_1).

##### Assessment Blueprints

The CSA blueprints describe each of the Spanish reading/language arts domains, including reading, writing mechanics, and listening for all tested grade levels and the high school grade band and how that content domain is assessed through the testable standards (CDE, 2017). Each test is described by a single blueprint. The degree to which test forms administered in 2021–22 meet the blueprint is provided in [*Chapter 3: Item Development and Review*](#_Item_Development_and_1) and [*Chapter 4: Test Assembly*](#_Test_Assembly_1) as well as in [appendix 4.A](#_Appendix_4.A_Test).

##### Form Assembly Process

Once items are developed and field-tested, ETS selects all CSA items to conform with the SBE-approved CSA content standards and test blueprint. The content standards, blueprint, and test specifications were used as the basis for choosing items for the CSA. Refer to [*Chapter 4: Test Assembly*](#_Test_Assembly) for information on the test assembly process.

#### Evidence Based on the 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 students” (AERA et al., 2014, p. 12). This type of evidence generally includes documentation of activities such as

* systematic observations of test response behavior,
* analysis of student item-response-time data, and
* evaluation of the reasoning processes students employ when solving test items (Embretson, 1983; Messick, 1989).

This type of evidence is used to confirm that the CSA is measuring the cognitive skills that are intended as the objects of measurement, and that students are using these targeted skills to respond to the items.

##### Analysis of Testing Time

Testing time for each test administration can be evaluated for consistency by examining the expected response processes for the items presented to students. The length of time it takes students to complete an assessment is collected and analyzed to build a profile describing what a typical testing event looks like for each grade level or the high school grade band. 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 CSA grade-level assessments are untimed.

Students with no item response and those who did not answer at least 10 items were removed from these analyses. The remaining testing population was partitioned into quartiles based on raw scores. These quartile groupings are not the same as the reporting ranges.

Descriptive statistics of the time required to complete the total test are computed for each of the four quartile groups by grade level. Some cases of extremely long testing time may be attributed to a test’s not being closed down properly. Note that the overall raw scores are used to get the raw score intervals, but the total testing time will be based on the whole test, including the field test items taken by a student.

Table 8.C.1 through table 8.C.7 in [appendix 8.C](#_Appendix_8.C:_Testing) present total testing time and raw score interval information at each student performance quartile level by grade level or the high school grade band. Overall, students who took the general forms at the lowest raw score interval have shorter testing times than students in the other quartile groups. The median total testing time generally increases as the quartile level increases from the lowest raw score interval to the highest raw score interval—that is, students who performed better on the CSA tended to spend more time on the test.

Note that the mean testing time should be interpreted with caution. Some cases of extremely long testing time may be attributed to students with special needs taking longer to complete a test or the test not being closed down properly. Therefore, mean testing times may be misleading.

The medians (fiftieth percentile) are more meaningful in the interpretation of the time comparisons because medians are less impacted by the extreme values than means. The removal of the 1 percent of the student data with the shortest testing times is a modest exclusion that leaves some students with very short durations in the results for each of the tests. Similarly, some very long durations are present in the data, which may indicate errors such as the failure to close a testing session. Therefore, the median is a better statistic than the mean for evaluating testing time information.

#### Evidence Based on the Internal Structure

Internal structure evidence evaluates the strength or salience of the major dimensions underlying an assessment. For the CSA, it is assumed that a single construct underlies the total scores obtained on each assessment. Evidence to support this assumption in a typical testing year can be gathered from the results of item analyses, DIF analyses, evaluations of internal consistency, and studies of reliability.

##### Classical Statistics

Polyserial correlations calculated for the items in an assessment show the degree to which the items discriminate between students with low and high scores on an assessment. To the degree that the correlations are high, evidence that the items assess the same construct is provided. As shown in table 8.5, the mean polyserial correlation was between 0.39 and 0.44. The polyserial correlations for the individual items, as well as flag status, in the CSA are presented in table 8.A.3 through table 8.A.9 in [appendix 8.A](#_Appendix_8.A:_Classical).

Also relevant to the validity of a score interpretation are the ranges of item difficulty for the items on which a test score will be based. The finding—that items have difficulties that span the range of student ability—provides evidence that students at all levels of ability are adequately measured by the items. Information on average item *p*-values is given in table 8.4; the data in table 8.4 indicates that these assessments had average *p*-values ranging from 0.44 to 0.50. Individual item *p*-values are also presented in table 8.A.3 through table 8.A.9.

##### Differential Item Functioning

DIF analyses are used to assess differences in the item performance of groups of students who differ in their demographic characteristics. Refer to section [*8.3 Differential Item Functioning Analyses*](#_Differential_Item_Functioning) for a description of the DIF analyses and table 8.10, where the results of the DIF analyses are reported.

##### Overall Reliability Estimates

The reliability analysis, including reliability at the student level and group level, is described in section [*8.6 Reliability Analyses*](#_Reliability_Analyses). The results of reliability analyses on the overall raw score for each CSA form are presented in table 8.14. The results indicate that the reliability estimates for all general and accommodated forms are satisfactory, ranging from 0.82 to 0.86 for the general forms and 0.77 to 0.85 for the accommodated forms.

##### Student Groups Reliability Estimates

The reliabilities were also examined for various student groups that differ in their demographic characteristics within the student population. The student groups considered were gender, economic status, provision of special services, length of attendance in US schools, whether students received instruction in Spanish, and English language fluency levels. Student groups for accommodated forms were not analyzed because of the extremely small sample size for these forms.

Across student groups, reliability coefficients were generally higher than 0.80 except for some groups of students who received special education services, groups not receiving instruction in Spanish in grades three and seven, and EL student groups in grades four through eight. The reliability was also lower because of a lack of variation in performance caused by small group size and homogenous group members. Refer to subsection [*8.6.4 Student Group Reliabilities and Standard Errors of Measurement*](#_Student_Group_Reliabilities) for details.

Reliability estimates and SEM information for the total test scores are reported for each student group in table 8.D.1 through table 8.D.6 in [appendix 8.D](#_Appendix_8.D:_Reliability_1).

##### Reliability of Performance Classifications

The methodology used for estimating the reliability of classification decisions is described with the decision classification analyses in subsection [*8.6.6 Decision Classification Analyses*](#_Decision_Classification_Analyses). The overall decision accuracy is greater than 0.80 for all seven assessments. The overall decision consistency is relatively lower, with the lowest being 0.74 for grade six and the highest being 0.83 for grade three. The results of these analyses are presented in table 8.E.1 through table 8.E.14 in [appendix 8.E](#_Appendix_8.E:_Analyses).

### 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.

Browne, M. W., & Cudek, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J.S. Long (Eds.) *Testing structural equation models* (pp. 136–62)*.* Newbury Park, CA: Sage Publications.

Cai, L. (2017). FlexMIRT®: *Flexible multilevel, multidimensional item analysis and test scoring* (Version 3.5) [computer software]. Chapel Hill, NC: Vector Psychometric Group.

California Department of Education. (2017). *California Spanish Assessment blueprint.* Sacramento, CA: California Department of Education.

California Department of Education. (2022). *CAASPP online test administration manual.* Sacramento, CA: California Department of Education.

Crocker, L. M., Miller, D., & Franks, E. A. (1989). Quantitative methods for assessing the fit between test and curriculum. *Applied Measurement in Education, 2,* 179–94.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

Dorans, N. J. (2013). ETS contributions to the quantitative assessment of item, test, and score fairness. *ETS Research Report Series*, i–38.

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

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

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

Embretson (Whitley), S. (1983). Construct validity: construct representation versus nomothetic span. *Psychological Bulletin, 93,* 179–97.

Feldt, L. S. & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd edition) (pp. 105–46). New York: Macmillan.

Feldt, L. S. & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd edition) (pp. 105–46). New York: Macmillan.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist,* 18. 519–32.

Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65–110). Washington, DC: American Council on Education and National Council on Measurement in Education.

Holland, P. W., & Thayer, D. T. (1985). *An alternative definition of the ETS delta scale of item difficulty* (Research Report 85–43). Princeton, NJ: Educational Testing Service.

Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Erlbaum.

Huynh, H. (2000, June). *Guidelines for Rasch linking for PACT*. Memorandum to Paul Sandifer on June 18, 2000. Columbia, SC: Available from Author.

Huynh, H., & Rawls, A. (2009). A comparison between robust z and 0.3-logit difference procedures in assessing stability of linking items for the Rasch model. In Everett V. Smith Jr. & Greg E. Stone (Eds.) *Applications of Rasch Measurement in Criterion-Reference Testing: Practice analysis to Score Reporting.* Maple Grove, MN: JAM Press.

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

Kim, S., & Kolen M. J. (2004). STUIRT: A Computer Program for Scale Transformation under Unidimensional Item Response Theory Models. University of Iowa. Version 1.0.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, linking, and scaling: Methods and practices* (2nd ed.). New York, NY: Springer-Verlag.

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

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

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

Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessments, and instruction. *Review of Educational Research, 4,* 1332–61.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–‍74.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed.). Washington, DC: American Council on Education.

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.

San Diego County Office of Education. (n.d.) *CCSS en Español: Common Core State Standards*. San Diego County Office of Education website.

Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment, 5,* 299–321.

Stocking, M. L. (1996). An alternative method for scoring adaptive tests. *Journal of Educational and Behavioral Statistics, 21*, 365–89.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7(2), 201–10.

Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). Descriptive and inferential procedures for assessing differential item functioning in polytomous items. *Applied Measurement in Education, 10*(4), 321–44.

### Accessibility Information

#### Alternative Text for Equation 8.1

P value sub dich equals the fraction with the numerator the sum of X sub ij and the denominator N sub I end fraction.

#### Alternative Text for Equation 8.2

P value sub poly equals the fraction with the numerator X sub ij and the denominator N sub i times Max of X sub I end fraction.

#### Alternative Text for Equation 8.3

r sub polyreg equals the fraction beta sub hat times S tot divided by the square root of Beta sub hat squared times s sub tot squared plus 1.

#### Alternative Text for Equation 8.4

Alpha sub MH equals the numerator open parenthesis the sum sub m of R sub rm times W sub fm divided by N sub tm close parenthesis divided by the denominator open parenthesis the sum sub m of R sub fm times W sub rm divided by N sub tm closed parenthesis.

#### Alternative Text for Equation 8.5

MH D-DIF equals negative 2.35 times the natural logarithm open bracket alpha sub MH close bracket.

#### Alternative Text for Equation 8.6

SMD equals the fraction with numerator the sum from m equals 1 to M of N sub fm times E sub 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 8.7

P sub ih of theta sub j equals the numerator exp open parenthesis the sum from v equals 1 to h of Da sub i open parenthesis theta sub j minus b sub I plus d sub iv close parenthesis close parenthesis divided by the denominator open parenthesis 1 plus the sum from c equals 1 to n sub I exp open parenthesis the sum from v equals 1 to c of Da sub I open parenthesis theta sub j minus b sub I plus d sub iv close parenthesis close parenthesis close parenthesis, if score h equals 1, 2, …, n sub i.

P sub ih of theta sub j equals 1 divided by the denominator open parenthesis 1 plus the sum from c equals 1 to n sub I exp open parenthesis the sum from v equals 1 to c of Da sub I open parenthesis theta sub j minus b sub I plus d sub iv close parenthesis close parenthesis close parenthesis, if score h equals 0.

#### Alternative Text for Equation 8.8

z equals the numerator open absolute symbol, D subtracts Md sub D, close absolute symbol, divided by the denominator of 0.74 times IQR.

#### Alternative Text for Equation 8.9

d-squared equals the sum across k quadrature points the squared difference in the probability of a correct response-to based on the new parameters and bank parameters multiplied by weights.

#### Alternative Text for Equation 8.10

Epsilon of theta equals the sum from i equals 1 to ndich of P sub i of theta plus the sum from j equals 1 to npoly over each sum of x equals 1 to m of s sub xj times P sub xj of theta.

#### Alternative Text for Equation 8.11

Scale score equals intercept a plus slope times theta.

#### Alternative Text for Equation 8.12

Slope *a* equals sigma sub scale score divided by sigma sub theta.

#### Alternative Text for Equation 8.13

Intercept *b* equals scale score at threshold 3 subtract sigma sub scale score divided by sigma sub theta times theta sub threshold 3.

#### Alternative Text for Equation 8.14

Scale score equals scale score at threshold 3 plus sigma sub scale score divided by sigma sub theta times open bracket theta minus theta sub threshold 3 close bracket.

#### Alternative Text for Equation 8.15

Alpha hat equals the numerator k divided by the denominator k minus 1 multiplied by open bracket 1 minus the numerator summation from i equals 1 to k multiplied by sigma-hat squared sub i divided by the denominator sigma-hat squared sub x close bracket.

#### Alternative Text for Equation 8.16

Alpha hat sub c equals 1 minus fraction with numerator sum of j of w squared sub j times Sigma hat squared sub j times open parenthesis 1 minus alpha hat sub j close parenthesis and denominator Sigma hat squared sub c.

#### Alternative Text for Equation 8.17

SEM equals S sub X times square root of 1 minus alpha hat sub c.

#### Alternative Text for Equation 8.18

SE of theta sub j equals 1 divided by the square root of I of theta sub j.

#### Alternative Text for Equation 8.19

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 8.20

I sub i of theta sub j equals open bracket s sub i2 of theta sub j minus s sub i squared of theta sub j.

#### Alternative Text for Equation 8.21

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 8.22

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 8.23

CSEM of SS equals 1 times a divided by the square root of I of theta hat.

### Appendix 8.A: Classical Item Analyses

Table 8.A.1 Item Difficulty Distributions by Domain

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **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** | **Mean *p-*‍value** | **Minimum *p-*value** | **Maximum *p-*value** |
| Grade 3 | Reading | 0 | 14 | 10 | 0 | 0 | 24 | 0.39 | 0.21 | 0.59 |
| Grade 3 | Writing Mechanics | 0 | 5 | 9 | 2 | 0 | 16 | 0.45 | 0.20 | 0.61 |
| Grade 3 | Listening | 0 | 1 | 9 | 2 | 0 | 12 | 0.52 | 0.30 | 0.79 |
| Grade 4 | Reading | 0 | 11 | 12 | 1 | 0 | 24 | 0.41 | 0.22 | 0.60 |
| Grade 4 | Writing Mechanics | 0 | 6 | 7 | 3 | 0 | 16 | 0.47 | 0.29 | 0.71 |
| Grade 4 | Listening | 1 | 2 | 5 | 4 | 0 | 12 | 0.51 | 0.20 | 0.73 |
| Grade 5 | Reading | 2 | 9 | 12 | 1 | 0 | 24 | 0.41 | 0.19 | 0.79 |
| Grade 5 | Writing Mechanics | 0 | 7 | 8 | 1 | 0 | 16 | 0.42 | 0.21 | 0.62 |
| Grade 5 | Listening | 0 | 2 | 2 | 7 | 1 | 12 | 0.59 | 0.34 | 0.82 |
| Grade 6 | Reading | 0 | 5 | 16 | 3 | 0 | 24 | 0.47 | 0.21 | 0.67 |
| Grade 6 | Writing Mechanics | 0 | 4 | 6 | 5 | 1 | 16 | 0.53 | 0.26 | 0.81 |
| Grade 6 | Listening | 0 | 2 | 6 | 4 | 0 | 12 | 0.51 | 0.26 | 0.73 |
| Grade 7 | Reading | 1 | 11 | 11 | 1 | 0 | 24 | 0.39 | 0.04 | 0.61 |
| Grade 7 | Writing Mechanics | 0 | 5 | 7 | 4 | 0 | 16 | 0.48 | 0.31 | 0.67 |
| Grade 7 | Listening | 0 | 2 | 6 | 4 | 0 | 12 | 0.56 | 0.36 | 0.77 |
| Grade 8 | Reading | 0 | 8 | 15 | 1 | 0 | 24 | 0.43 | 0.25 | 0.63 |
| Grade 8 | Writing Mechanics | 0 | 5 | 7 | 4 | 0 | 16 | 0.49 | 0.27 | 0.78 |
| Grade 8 | Listening | 0 | 1 | 6 | 4 | 1 | 12 | 0.58 | 0.32 | 0.80 |
| High School | Reading | 0 | 9 | 14 | 1 | 0 | 24 | 0.42 | 0.24 | 0.61 |
| High School | Writing Mechanics | 1 | 7 | 5 | 3 | 0 | 16 | 0.41 | 0.16 | 0.69 |
| High School | Listening | 0 | 1 | 9 | 2 | 0 | 12 | 0.52 | 0.39 | 0.68 |

Table 8.A.2 Item-Total Correlation Distributions by Domain

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **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** | **Mean r** | **Minimum r** | **Maximum r** |
| Grade 3 | Reading | 0 | 1 | 4 | 5 | 10 | 4 | 24 | 0.41 | 0.15 | 0.73 |
| Grade 3 | Writing Mechanics | 0 | 2 | 2 | 6 | 3 | 3 | 16 | 0.38 | 0.13 | 0.61 |
| Grade 3 | Listening | 0 | 0 | 1 | 1 | 2 | 8 | 12 | 0.51 | 0.26 | 0.66 |
| Grade 4 | Reading | 0 | 1 | 7 | 5 | 4 | 7 | 24 | 0.41 | 0.11 | 0.70 |
| Grade 4 | Writing Mechanics | 0 | 0 | 2 | 7 | 5 | 2 | 16 | 0.40 | 0.27 | 0.64 |
| Grade 4 | Listening | 0 | 0 | 1 | 3 | 3 | 5 | 12 | 0.48 | 0.23 | 0.68 |
| Grade 5 | Reading | 0 | 0 | 3 | 10 | 6 | 5 | 24 | 0.41 | 0.24 | 0.56 |
| Grade 5 | Writing Mechanics | 0 | 1 | 2 | 7 | 4 | 2 | 16 | 0.37 | 0.19 | 0.61 |
| Grade 5 | Listening | 0 | 0 | 3 | 1 | 3 | 5 | 12 | 0.49 | 0.21 | 0.74 |
| Grade 6 | Reading | 0 | 2 | 2 | 7 | 8 | 5 | 24 | 0.42 | 0.11 | 0.70 |
| Grade 6 | Writing Mechanics | 0 | 1 | 2 | 2 | 4 | 7 | 16 | 0.44 | 0.17 | 0.59 |
| Grade 6 | Listening | 0 | 0 | 2 | 1 | 4 | 5 | 12 | 0.46 | 0.27 | 0.57 |
| Grade 7 | Reading | 0 | 2 | 6 | 8 | 3 | 5 | 24 | 0.36 | 0.09 | 0.67 |
| Grade 7 | Writing Mechanics | 0 | 0 | 5 | 3 | 5 | 3 | 16 | 0.38 | 0.20 | 0.60 |
| Grade 7 | Listening | 0 | 0 | 0 | 4 | 6 | 2 | 12 | 0.44 | 0.31 | 0.67 |
| Grade 8 | Reading | 0 | 1 | 4 | 8 | 5 | 6 | 24 | 0.41 | 0.18 | 0.65 |
| Grade 8 | Writing Mechanics | 0 | 1 | 0 | 6 | 5 | 4 | 16 | 0.43 | 0.15 | 0.65 |
| Grade 8 | Listening | 0 | 0 | 1 | 4 | 3 | 4 | 12 | 0.44 | 0.29 | 0.64 |
| High School | Reading | 1 | 0 | 3 | 5 | 11 | 4 | 24 | 0.39 | -0.14 | 0.55 |
| High School | Writing Mechanics | 0 | 0 | 6 | 3 | 4 | 3 | 16 | 0.39 | 0.23 | 0.64 |
| High School | Listening | 0 | 0 | 1 | 3 | 5 | 3 | 12 | 0.41 | 0.25 | 0.55 |

**Notes:**

What follows are the values that can appear in the *Flag* column in table 8.A.3 through table 8.A.9:

* A = low average item score
* Dichotomous
* Item with two choices ≤ less than 0.50
* Item with three choices ≤ less than 0.30
* Item with four choices ≤ less than 0.20
* Polytomous = < 0.20
* D = proportionally more higher-ability test-takers (top 20%) selected a distractor over the key
* H = high average item score (> 0.95)
* L = underrepresented score category (< 3%)
* O = high percent of omits (> 5%)
* P = positive correlation with the criterion score for one or more of the distractors
* Rpoly = low correlation with criterion (< 0.20)

Table 8.A.3 Item Difficulty, Item-Total Correlation, and Flag Status by Item for Grade Three

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 1 | 0.44 | 0.39 | [no flag] | 1 | MC |
| 2 | 0.49 | 0.54 | [no flag] | 1 | MC |
| 3 | 0.43 | 0.42 | [no flag] | 1 | MC |
| 4 | 0.54 | 0.36 | [no flag] | 1 | InlineChoices |
| 5 | 0.50 | 0.52 | [no flag] | 2 | MC |
| 6 | 0.55 | 0.63 | [no flag] | 2 | InlineChoices |
| 7 | 0.48 | 0.62 | [no flag] | 1 | MC |
| 8 | 0.47 | 0.43 | [no flag] | 1 | MC |
| 9 | 0.33 | 0.43 | [no flag] | 1 | MC |
| 10 | 0.49 | 0.27 | [no flag] | 1 | InlineChoices |
| 11 | 0.45 | 0.40 | [no flag] | 1 | MC |
| 12 | 0.41 | 0.40 | [no flag] | 1 | MC |
| 13 | 0.35 | 0.53 | [no flag] | 1 | MC |
| 14 | 0.59 | 0.48 | [no flag] | 2 | MC |
| 15 | 0.43 | 0.26 | [no flag] | 1 | MC |
| 16 | 0.35 | 0.34 | [no flag] | 1 | MC |
| 17 | 0.27 | 0.41 | [no flag] | 1 | MC |
| 18 | 0.33 | 0.15 | P R | 1 | MC |

Table 8.A.3 *(continuation)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 19 | 0.53 | 0.36 | [no flag] | 1 | MC |
| 20 | 0.20 | 0.27 | [no flag] | 1 | MC |
| 21 | 0.29 | 0.13 | R | 1 | InlineChoices |
| 22 | 0.46 | 0.50 | [no flag] | 1 | MC |
| 23 | 0.47 | 0.37 | [no flag] | 1 | InlineChoices |
| 24 | 0.49 | 0.40 | [no flag] | 1 | InlineChoices |
| 25 | 0.27 | 0.61 | [no flag] | 1 | MC |
| 26 | 0.33 | 0.22 | [no flag] | 1 | MC |
| 27 | 0.58 | 0.50 | [no flag] | 2 | Composite |
| 28 | 0.30 | 0.26 | [no flag] | 1 | MC |
| 29 | 0.56 | 0.62 | [no flag] | 2 | InlineChoices |
| 30 | 0.50 | 0.56 | [no flag] | 1 | MC |
| 31 | 0.79 | 0.57 | [no flag] | 2 | Zones |
| 32 | 0.61 | 0.66 | [no flag] | 1 | InlineChoices |
| 33 | 0.61 | 0.50 | [no flag] | 2 | MC |
| 34 | 0.53 | 0.39 | [no flag] | 1 | MC |
| 35 | 0.48 | 0.50 | [no flag] | 2 | InlineChoices |
| 36 | 0.36 | 0.36 | [no flag] | 2 | Composite |
| 37 | 0.57 | 0.16 | R | 2 | InlineChoices |
| 38 | 0.60 | 0.39 | [no flag] | 2 | MC |
| 39 | 0.21 | 0.45 | [no flag] | 2 | Match |
| 40 | 0.53 | 0.40 | [no flag] | 1 | MC |
| 41 | 0.47 | 0.27 | [no flag] | 1 | InlineChoices |
| 42 | 0.44 | 0.22 | [no flag] | 1 | MC |
| 43 | 0.34 | 0.38 | [no flag] | 1 | MC |
| 44 | 0.32 | 0.48 | [no flag] | 1 | MC |
| 45 | 0.47 | 0.38 | [no flag] | 2 | MC |
| 46 | 0.35 | 0.51 | [no flag] | 1 | Match |
| 47 | 0.40 | 0.40 | D P | 2 | MC |
| 48 | 0.23 | 0.49 | [no flag] | 1 | MC |
| 49 | 0.40 | 0.46 | [no flag] | 1 | InlineChoices |
| 50 | 0.52 | 0.49 | [no flag] | 1 | MC |
| 51 | 0.39 | 0.73 | [no flag] | 1 | MC |
| 52 | 0.35 | 0.54 | [no flag] | 1 | MC |

Table 8.A.4 Item Difficulty and Item-Total Correlation by Item for Grade Four

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 1 | 0.37 | 0.40 | [no flag] | 1 | MC |
| 2 | 0.58 | 0.66 | [no flag] | 1 | MC |
| 3 | 0.70 | 0.47 | [no flag] | 1 | InlineChoices |
| 4 | 0.53 | 0.32 | D | 2 | Match |
| 5 | 0.42 | 0.45 | [no flag] | 1 | MC |
| 6 | 0.20 | 0.23 | A D | 1 | MC |
| 7 | 0.39 | 0.40 | [no flag] | 1 | MC |
| 8 | 0.29 | 0.45 | [no flag] | 1 | Grid |
| 9 | 0.36 | 0.46 | [no flag] | 1 | MC |
| 10 | 0.52 | 0.37 | D | 2 | InlineChoices |
| 11 | 0.61 | 0.42 | [no flag] | 1 | InlineChoices |
| 12 | 0.55 | 0.39 | [no flag] | 1 | MC |
| 13 | 0.46 | 0.30 | [no flag] | 2 | InlineChoices |
| 14 | 0.71 | 0.64 | [no flag] | 1 | InlineChoices |
| 15 | 0.35 | 0.51 | [no flag] | 1 | MC |
| 16 | 0.52 | 0.29 | [no flag] | 2 | MC |
| 17 | 0.35 | 0.26 | [no flag] | 1 | MC |
| 18 | 0.60 | 0.30 | P | 2 | MC |
| 19 | 0.24 | 0.42 | [no flag] | 1 | MC |
| 20 | 0.29 | 0.29 | [no flag] | 1 | InlineChoices |
| 21 | 0.53 | 0.25 | P | 2 | MC |
| 22 | 0.42 | 0.36 | [no flag] | 1 | MC |
| 23 | 0.31 | 0.36 | [no flag] | 1 | MC |
| 24 | 0.44 | 0.63 | [no flag] | 1 | MC |
| 25 | 0.47 | 0.70 | [no flag] | 1 | MC |
| 26 | 0.27 | 0.11 | R | 1 | MC |
| 27 | 0.22 | 0.20 | D | 1 | InlineChoices |
| 28 | 0.58 | 0.36 | [no flag] | 2 | InlineChoices |
| 29 | 0.47 | 0.62 | [no flag] | 1 | MC |
| 30 | 0.30 | 0.39 | A | 1 | Zones |
| 31 | 0.73 | 0.46 | [no flag] | 2 | Match |
| 32 | 0.73 | 0.54 | [no flag] | 2 | Zones |
| 33 | 0.41 | 0.58 | [no flag] | 1 | MC |
| 34 | 0.71 | 0.68 | [no flag] | 1 | MC |
| 35 | 0.48 | 0.27 | [no flag] | 1 | MC |
| 36 | 0.33 | 0.62 | [no flag] | 2 | Composite |
| 37 | 0.32 | 0.64 | [no flag] | 1 | MC |
| 38 | 0.43 | 0.65 | [no flag] | 1 | MC |

Table 8.A.4 *(continuation)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 39 | 0.57 | 0.29 | P | 2 | MC |
| 40 | 0.35 | 0.50 | [no flag] | 1 | InlineChoices |
| 41 | 0.37 | 0.33 | [no flag] | 1 | MC |
| 42 | 0.43 | 0.28 | [no flag] | 2 | MC |
| 43 | 0.39 | 0.36 | [no flag] | 2 | Composite |
| 44 | 0.47 | 0.41 | [no flag] | 1 | InlineChoices |
| 45 | 0.40 | 0.31 | [no flag] | 1 | MC |
| 46 | 0.41 | 0.36 | [no flag] | 1 | MC |
| 47 | 0.68 | 0.44 | [no flag] | 2 | MC |
| 48 | 0.42 | 0.34 | [no flag] | 1 | MC |
| 49 | 0.55 | 0.54 | [no flag] | 1 | MC |
| 50 | 0.50 | 0.45 | [no flag] | 1 | MC |
| 51 | 0.47 | 0.41 | [no flag] | 1 | MC |
| 52 | 0.39 | 0.44 | [no flag] | 1 | MC |

Table 8.A.5 Item Difficulty and Item-Total Correlation by Item for Grade Five

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 1 | 0.46 | 0.37 | D | 2 | InlineChoices |
| 2 | 0.62 | 0.31 | [no flag] | 1 | InlineChoices |
| 3 | 0.71 | 0.58 | [no flag] | 1 | InlineChoices |
| 4 | 0.66 | 0.65 | [no flag] | 1 | MC |
| 5 | 0.82 | 0.35 | [no flag] | 1 | Zones |
| 6 | 0.69 | 0.62 | [no flag] | 1 | MC |
| 7 | 0.66 | 0.48 | [no flag] | 2 | MC |
| 8 | 0.34 | 0.21 | [no flag] | 1 | InlineChoices |
| 9 | 0.54 | 0.28 | [no flag] | 1 | InlineChoices |
| 10 | 0.29 | 0.50 | [no flag] | 1 | MC |
| 11 | 0.52 | 0.36 | D P | 2 | MC |
| 12 | 0.43 | 0.46 | [no flag] | 1 | MC |
| 13 | 0.43 | 0.48 | [no flag] | 1 | MC |
| 14 | 0.19 | 0.33 | A P | 1 | MC |
| 15 | 0.32 | 0.35 | [no flag] | 1 | InlineChoices |
| 16 | 0.38 | 0.29 | [no flag] | 1 | MC |
| 17 | 0.21 | 0.31 | D | 1 | MC |
| 18 | 0.46 | 0.32 | [no flag] | 2 | InlineChoices |
| 19 | 0.43 | 0.41 | [no flag] | 2 | InlineChoices |
| 20 | 0.36 | 0.19 | R | 1 | MC |
| 21 | 0.31 | 0.41 | [no flag] | 1 | MC |
| 22 | 0.37 | 0.33 | [no flag] | 1 | MC |
| 23 | 0.47 | 0.41 | [no flag] | 1 | MC |
| 24 | 0.52 | 0.29 | [no flag] | 1 | MC |
| 25 | 0.36 | 0.56 | [no flag] | 2 | MC |
| 26 | 0.36 | 0.31 | [no flag] | 1 | MC |
| 27 | 0.49 | 0.47 | [no flag] | 1 | MC |
| 28 | 0.46 | 0.61 | [no flag] | 1 | MC |
| 29 | 0.30 | 0.33 | [no flag] | 1 | MC |
| 30 | 0.42 | 0.43 | [no flag] | 2 | InlineChoices |
| 31 | 0.61 | 0.50 | [no flag] | 1 | MC |
| 32 | 0.58 | 0.30 | P | 2 | MC |
| 33 | 0.67 | 0.74 | [no flag] | 1 | MC |
| 34 | 0.62 | 0.68 | [no flag] | 1 | MC |
| 35 | 0.35 | 0.29 | [no flag] | 1 | MC |
| 36 | 0.35 | 0.31 | [no flag] | 1 | MC |
| 37 | 0.39 | 0.34 | [no flag] | 1 | MC |
| 38 | 0.42 | 0.43 | [no flag] | 1 | MC |

Table 8.A.5 *(continuation)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 39 | 0.49 | 0.40 | [no flag] | 1 | Match |
| 40 | 0.46 | 0.52 | [no flag] | 1 | InlineChoices |
| 41 | 0.28 | 0.40 | [no flag] | 1 | MC |
| 42 | 0.25 | 0.24 | D | 1 | MC |
| 43 | 0.53 | 0.36 | [no flag] | 1 | InlineChoices |
| 44 | 0.35 | 0.49 | [no flag] | 2 | InlineChoices |
| 45 | 0.59 | 0.40 | [no flag] | 2 | MC |
| 46 | 0.39 | 0.47 | [no flag] | 1 | MC |
| 47 | 0.44 | 0.37 | [no flag] | 1 | MC |
| 48 | 0.49 | 0.56 | [no flag] | 1 | MC |
| 49 | 0.47 | 0.54 | [no flag] | 1 | MC |
| 50 | 0.43 | 0.27 | [no flag] | 2 | InlineChoices |
| 51 | 0.20 | 0.52 | A | 1 | MC |
| 52 | 0.79 | 0.52 | [no flag] | 2 | InlineChoices |

Table 8.A.6 Item Difficulty and Item-Total Correlation by Item for Grade Six

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 1 | 0.43 | 0.23 | [no flag] | 1 | InlineChoices |
| 2 | 0.60 | 0.27 | [no flag] | 1 | MC |
| 3 | 0.54 | 0.50 | [no flag] | 2 | MC |
| 4 | 0.31 | 0.48 | [no flag] | 1 | Grid |
| 5 | 0.59 | 0.56 | [no flag] | 1 | MC |
| 6 | 0.40 | 0.46 | [no flag] | 1 | MC |
| 7 | 0.65 | 0.57 | [no flag] | 1 | InlineChoices |
| 8 | 0.73 | 0.54 | [no flag] | 2 | MC |
| 9 | 0.59 | 0.43 | [no flag] | 1 | MC |
| 10 | 0.21 | 0.40 | [no flag] | 1 | InlineChoices |
| 11 | 0.42 | 0.48 | [no flag] | 2 | Match |
| 12 | 0.48 | 0.20 | R | 1 | MC |
| 13 | 0.65 | 0.47 | [no flag] | 2 | MC |
| 14 | 0.72 | 0.50 | [no flag] | 1 | InlineChoices |
| 15 | 0.52 | 0.36 | [no flag] | 1 | MC |
| 16 | 0.81 | 0.58 | H | 2 | InlineChoices |
| 17 | 0.38 | 0.57 | [no flag] | 1 | MC |
| 18 | 0.37 | 0.43 | [no flag] | 1 | MC |
| 19 | 0.48 | 0.59 | [no flag] | 1 | MC |
| 20 | 0.72 | 0.35 | [no flag] | 2 | InlineChoices |
| 21 | 0.46 | 0.50 | [no flag] | 1 | MC |
| 22 | 0.44 | 0.51 | [no flag] | 1 | MC |
| 23 | 0.45 | 0.45 | [no flag] | 1 | MC |
| 24 | 0.26 | 0.27 | [no flag] | 1 | MC |
| 25 | 0.78 | 0.57 | [no flag] | 1 | MC |
| 26 | 0.43 | 0.26 | [no flag] | 1 | MC |
| 27 | 0.36 | 0.17 | R | 1 | MC |
| 28 | 0.68 | 0.55 | [no flag] | 1 | MC |
| 29 | 0.49 | 0.44 | [no flag] | 1 | MC |
| 30 | 0.63 | 0.35 | [no flag] | 2 | InlineChoices |
| 31 | 0.26 | 0.51 | [no flag] | 1 | MC |
| 32 | 0.52 | 0.50 | [no flag] | 1 | MC |
| 33 | 0.51 | 0.48 | [no flag] | 1 | MC |
| 34 | 0.41 | 0.37 | [no flag] | 1 | Match |
| 35 | 0.67 | 0.48 | [no flag] | 2 | MC |
| 36 | 0.60 | 0.52 | [no flag] | 2 | MC |
| 37 | 0.62 | 0.65 | [no flag] | 1 | InlineChoices |
| 38 | 0.63 | 0.60 | [no flag] | 1 | MC |

Table 8.A.6 *(continuation)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 39 | 0.51 | 0.34 | P | 2 | MC |
| 40 | 0.30 | 0.55 | [no flag] | 1 | MC |
| 41 | 0.55 | 0.45 | [no flag] | 1 | MC |
| 42 | 0.49 | 0.43 | [no flag] | 1 | MC |
| 43 | 0.48 | 0.35 | [no flag] | 1 | MC |
| 44 | 0.57 | 0.70 | [no flag] | 1 | MC |
| 45 | 0.43 | 0.37 | [no flag] | 1 | MC |
| 46 | 0.30 | 0.27 | [no flag] | 1 | MC |
| 47 | 0.42 | 0.29 | D P | 2 | InlineChoices |
| 48 | 0.46 | 0.11 | D P R | 2 | MC |
| 49 | 0.54 | 0.49 | [no flag] | 1 | MC |
| 50 | 0.36 | 0.38 | [no flag] | 1 | MC |
| 51 | 0.43 | 0.36 | [no flag] | 2 | Composite |
| 52 | 0.39 | 0.42 | [no flag] | 1 | MC |

Table 8.A.7 Item Difficulty and Item-Total Correlation by Item for Grade Seven

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 1 | 0.61 | 0.26 | [no flag] | 1 | MC |
| 2 | 0.37 | 0.32 | [no flag] | 2 | InlineChoices |
| 3 | 0.56 | 0.40 | [no flag] | 2 | MC |
| 4 | 0.77 | 0.67 | [no flag] | 1 | MC |
| 5 | 0.54 | 0.43 | [no flag] | 1 | MC |
| 6 | 0.36 | 0.31 | [no flag] | 1 | MC |
| 7 | 0.65 | 0.45 | [no flag] | 1 | MC |
| 8 | 0.38 | 0.20 | [no flag] | 1 | InlineChoices |
| 9 | 0.46 | 0.59 | [no flag] | 1 | Grid |
| 10 | 0.38 | 0.22 | P | 1 | MC |
| 11 | 0.50 | 0.33 | P | 2 | MC |
| 12 | 0.41 | 0.36 | [no flag] | 1 | MC |
| 13 | 0.30 | 0.33 | [no flag] | 1 | MC |
| 14 | 0.35 | 0.34 | [no flag] | 1 | MC |
| 15 | 0.47 | 0.40 | [no flag] | 1 | InlineChoices |
| 16 | 0.54 | 0.43 | [no flag] | 2 | MC |
| 17 | 0.50 | 0.44 | [no flag] | 1 | InlineChoices |
| 18 | 0.50 | 0.60 | [no flag] | 1 | MC |
| 19 | 0.36 | 0.24 | [no flag] | 1 | MC |
| 20 | 0.53 | 0.52 | [no flag] | 1 | MC |
| 21 | 0.31 | 0.25 | [no flag] | 1 | MC |
| 22 | 0.34 | 0.51 | [no flag] | 1 | MC |
| 23 | 0.54 | 0.44 | [no flag] | 1 | MC |
| 24 | 0.44 | 0.44 | [no flag] | 1 | MC |
| 25 | 0.56 | 0.37 | [no flag] | 1 | MC |
| 26 | 0.57 | 0.54 | [no flag] | 1 | MC |
| 27 | 0.65 | 0.48 | [no flag] | 2 | MC |
| 28 | 0.66 | 0.46 | [no flag] | 2 | MC |
| 29 | 0.36 | 0.20 | [no flag] | 1 | MC |
| 30 | 0.48 | 0.34 | [no flag] | 1 | MC |
| 31 | 0.42 | 0.43 | [no flag] | 1 | InlineChoices |
| 32 | 0.67 | 0.31 | P | 2 | MC |
| 33 | 0.41 | 0.27 | [no flag] | 1 | InlineChoices |
| 34 | 0.65 | 0.46 | [no flag] | 2 | InlineChoices |
| 35 | 0.60 | 0.45 | [no flag] | 2 | MC |
| 36 | 0.34 | 0.40 | [no flag] | 1 | MC |
| 37 | 0.23 | 0.52 | [no flag] | 1 | MC |
| 38 | 0.42 | 0.47 | [no flag] | 1 | MC |

Table 8.A.7 *(continuation)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 39 | 0.37 | 0.24 | P | 2 | MC |
| 40 | 0.38 | 0.39 | [no flag] | 1 | MC |
| 41 | 0.54 | 0.22 | [no flag] | 1 | InlineChoices |
| 42 | 0.35 | 0.15 | D P R | 1 | MC |
| 43 | 0.04 | 0.25 | A D | 1 | Grid |
| 44 | 0.25 | 0.51 | [no flag] | 1 | MC |
| 45 | 0.61 | 0.67 | [no flag] | 2 | InlineChoices |
| 46 | 0.46 | 0.57 | [no flag] | 1 | Match |
| 47 | 0.50 | 0.36 | [no flag] | 1 | MC |
| 48 | 0.26 | 0.39 | [no flag] | 1 | InlineChoices |
| 49 | 0.46 | 0.42 | P | 2 | MC |
| 50 | 0.51 | 0.27 | [no flag] | 2 | MC |
| 51 | 0.45 | 0.41 | [no flag] | 1 | MC |
| 52 | 0.37 | 0.09 | P R | 1 | InlineChoices |

Table 8.A.8 Item Difficulty and Item-Total Correlation by Item for Grade Eight

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 1 | 0.78 | 0.47 | [no flag] | 1 | MC |
| 2 | 0.64 | 0.55 | [no flag] | 2 | MC |
| 3 | 0.50 | 0.37 | [no flag] | 1 | MC |
| 4 | 0.80 | 0.64 | [no flag] | 1 | MC |
| 5 | 0.66 | 0.29 | [no flag] | 2 | MC |
| 6 | 0.59 | 0.35 | [no flag] | 1 | MC |
| 7 | 0.50 | 0.54 | [no flag] | 1 | MC |
| 8 | 0.48 | 0.46 | [no flag] | 1 | MC |
| 9 | 0.49 | 0.48 | [no flag] | 1 | MC |
| 10 | 0.63 | 0.48 | [no flag] | 2 | MC |
| 11 | 0.46 | 0.36 | [no flag] | 1 | MC |
| 12 | 0.43 | 0.34 | [no flag] | 1 | MC |
| 13 | 0.63 | 0.39 | [no flag] | 2 | InlineChoices |
| 14 | 0.39 | 0.34 | [no flag] | 1 | MC |
| 15 | 0.51 | 0.33 | [no flag] | 1 | MC |
| 16 | 0.29 | 0.18 | P R | 1 | MC |
| 17 | 0.40 | 0.38 | [no flag] | 1 | MC |
| 18 | 0.56 | 0.22 | P | 2 | MC |
| 19 | 0.45 | 0.25 | D P | 2 | InlineChoices |
| 20 | 0.32 | 0.32 | [no flag] | 1 | InlineChoices |
| 21 | 0.64 | 0.42 | [no flag] | 1 | MC |
| 22 | 0.56 | 0.35 | P | 2 | MC |
| 23 | 0.63 | 0.48 | [no flag] | 1 | InlineChoices |
| 24 | 0.59 | 0.50 | [no flag] | 1 | MC |
| 25 | 0.33 | 0.55 | [no flag] | 1 | MC |
| 26 | 0.38 | 0.26 | P | 1 | MC |
| 27 | 0.39 | 0.37 | [no flag] | 1 | MC |
| 28 | 0.63 | 0.33 | [no flag] | 2 | InlineChoices |
| 29 | 0.48 | 0.52 | [no flag] | 1 | MC |
| 30 | 0.33 | 0.48 | [no flag] | 1 | MC |
| 31 | 0.37 | 0.44 | [no flag] | 1 | Match |
| 32 | 0.38 | 0.15 | R | 1 | MC |
| 33 | 0.65 | 0.65 | [no flag] | 1 | Match |
| 34 | 0.27 | 0.37 | [no flag] | 1 | MC |
| 35 | 0.58 | 0.53 | [no flag] | 2 | MC |
| 36 | 0.41 | 0.63 | [no flag] | 1 | MC |
| 37 | 0.45 | 0.57 | [no flag] | 1 | MC |
| 38 | 0.48 | 0.49 | [no flag] | 1 | MC |

Table 8.A.8 *(continuation)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 39 | 0.50 | 0.63 | [no flag] | 1 | MC |
| 40 | 0.38 | 0.46 | [no flag] | 1 | MC |
| 41 | 0.58 | 0.59 | [no flag] | 1 | InlineChoices |
| 42 | 0.52 | 0.49 | [no flag] | 1 | MC |
| 43 | 0.44 | 0.31 | P | 2 | InlineChoices |
| 44 | 0.51 | 0.37 | P | 2 | Composite |
| 45 | 0.25 | 0.45 | P | 1 | Grid |
| 46 | 0.41 | 0.65 | [no flag] | 1 | MC |
| 47 | 0.55 | 0.33 | P | 2 | MC |
| 48 | 0.28 | 0.59 | [no flag] | 1 | MC |
| 49 | 0.44 | 0.44 | [no flag] | 1 | MC |
| 50 | 0.41 | 0.36 | [no flag] | 1 | MC |
| 51 | 0.43 | 0.35 | [no flag] | 1 | MC |
| 52 | 0.36 | 0.28 | [no flag] | 1 | MC |

Table 8.A.9 Item Difficulty and Item-Total Correlation by Item for High School

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 1 | 0.28 | 0.30 | [no flag] | 1 | MC |
| 2 | 0.68 | 0.31 | [no flag] | 2 | MC |
| 3 | 0.56 | 0.53 | [no flag] | 1 | MC |
| 4 | 0.51 | 0.40 | [no flag] | 1 | Zones |
| 5 | 0.44 | 0.25 | [no flag] | 1 | MC |
| 6 | 0.48 | 0.39 | [no flag] | 1 | MC |
| 7 | 0.63 | 0.40 | P | 2 | MC |
| 8 | 0.58 | 0.41 | [no flag] | 2 | MC |
| 9 | 0.51 | 0.37 | [no flag] | 2 | InlineChoices |
| 10 | 0.41 | 0.45 | [no flag] | 1 | InlineChoices |
| 11 | 0.48 | 0.52 | [no flag] | 1 | MC |
| 12 | 0.16 | 0.37 | A D | 1 | InlineChoices |
| 13 | 0.22 | 0.23 | D P | 1 | MC |
| 14 | 0.64 | 0.56 | [no flag] | 1 | MC |
| 15 | 0.57 | 0.28 | [no flag] | 1 | MC |
| 16 | 0.61 | 0.42 | [no flag] | 1 | MC |
| 17 | 0.39 | 0.36 | [no flag] | 1 | MC |
| 18 | 0.52 | 0.41 | [no flag] | 1 | MC |
| 19 | 0.53 | 0.43 | [no flag] | 1 | MC |
| 20 | 0.56 | 0.55 | [no flag] | 2 | MC |
| 21 | 0.49 | 0.42 | [no flag] | 1 | MC |
| 22 | 0.45 | 0.54 | [no flag] | 1 | MC |
| 23 | 0.45 | 0.33 | [no flag] | 1 | MC |
| 24 | 0.30 | 0.25 | [no flag] | 1 | MC |
| 25 | 0.32 | 0.51 | [no flag] | 1 | MC |
| 26 | 0.35 | 0.38 | [no flag] | 1 | MC |
| 27 | 0.24 | -0.14 | D P R | 1 | MC |
| 28 | 0.44 | 0.46 | [no flag] | 1 | MC |
| 29 | 0.59 | 0.51 | [no flag] | 2 | MC |
| 30 | 0.44 | 0.47 | [no flag] | 1 | InlineChoices |
| 31 | 0.47 | 0.45 | [no flag] | 1 | MC |
| 32 | 0.61 | 0.45 | [no flag] | 1 | InlineChoices |
| 33 | 0.37 | 0.62 | [no flag] | 1 | MC |
| 34 | 0.52 | 0.29 | [no flag] | 2 | InlineChoices |
| 35 | 0.31 | 0.24 | P | 1 | MC |
| 36 | 0.26 | 0.36 | [no flag] | 1 | InlineChoices |
| 37 | 0.26 | 0.36 | [no flag] | 1 | MC |
| 38 | 0.69 | 0.64 | [no flag] | 2 | MC |

Table 8.A.9 *(continuation)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item Sequence** | ***p*-value** | **r** | **Flag** | **Maximum Score Points** | **Item Type** |
| 39 | 0.44 | 0.49 | [no flag] | 1 | MC |
| 40 | 0.32 | 0.30 | [no flag] | 1 | MC |
| 41 | 0.44 | 0.26 | P | 2 | MC |
| 42 | 0.37 | 0.33 | P | 2 | MC |
| 43 | 0.24 | 0.24 | P | 1 | MC |
| 44 | 0.42 | 0.43 | [no flag] | 1 | InlineChoices |
| 45 | 0.41 | 0.47 | [no flag] | 1 | InlineChoices |
| 46 | 0.25 | 0.55 | [no flag] | 1 | MC |
| 47 | 0.45 | 0.43 | [no flag] | 1 | MC |
| 48 | 0.37 | 0.38 | [no flag] | 1 | MC |
| 49 | 0.46 | 0.43 | [no flag] | 1 | InlineChoices |
| 50 | 0.48 | 0.43 | [no flag] | 1 | MC |
| 51 | 0.45 | 0.45 | [no flag] | 1 | MC |
| 52 | 0.32 | 0.46 | [no flag] | 1 | MC |

### Appendix 8.B: Item Response Theory Analyses

Table 8.B.1 IRT Item Difficulty for Grade Three

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b*-parameter Standard Error** | ***d*-parameters** | ***d*-parameters Standard Error** |
| 1 | -0.0441 | 0.0241 | N/A | N/A |
| 2 | -0.2368 | 0.0244 | N/A | N/A |
| 3 | -0.0054 | 0.0247 | N/A | N/A |
| 4 | -0.4806 | 0.0239 | N/A | N/A |
| 5 | -0.2753 | 0.0202 | 1.1743, -1.1743 | 0.0209, 0.0209 |
| 6 | -0.5026 | 0.0181 | 0.2760, -0.2760 | 0.0219, 0.0219 |
| 7 | -0.2216 | 0.0249 | N/A | N/A |
| 8 | -0.1638 | 0.0242 | N/A | N/A |
| 9 | 0.5100 | 0.0258 | N/A | N/A |
| 10 | -0.2315 | 0.0240 | N/A | N/A |
| 11 | -0.0567 | 0.0244 | N/A | N/A |
| 12 | 0.1144 | 0.0245 | N/A | N/A |
| 13 | 0.3659 | 0.0254 | N/A | N/A |
| 14 | -0.7419 | 0.0194 | 0.8703, -0.8703 | 0.0213, 0.0213 |
| 15 | -0.0110 | 0.0241 | N/A | N/A |
| 16 | 0.3843 | 0.0250 | N/A | N/A |
| 17 | 0.7981 | 0.0266 | N/A | N/A |
| 18 | 0.4876 | 0.0252 | N/A | N/A |
| 19 | -0.4486 | 0.0240 | N/A | N/A |
| 20 | 1.2320 | 0.0289 | N/A | N/A |
| 21 | 0.6827 | 0.0261 | N/A | N/A |
| 22 | -0.1158 | 0.0242 | N/A | N/A |
| 23 | -0.1728 | 0.0242 | N/A | N/A |
| 24 | -0.2546 | 0.0240 | N/A | N/A |
| 25 | 0.8176 | 0.0281 | N/A | N/A |
| 26 | 0.5093 | 0.0252 | N/A | N/A |
| 27 | -0.6550 | 0.0184 | 0.5481, -0.5481 | 0.0212, 0.0212 |
| 28 | 0.6423 | 0.0259 | N/A | N/A |
| 29 | -0.5730 | 0.0185 | 0.5188, -0.5188 | 0.0215, 0.0215 |
| 30 | -0.2919 | 0.0245 | N/A | N/A |
| 31 | -1.9083 | 0.0297 | 1.0042, -1.0042 | 0.0325, 0.0325 |
| 32 | -0.7925 | 0.0249 | N/A | N/A |
| 33 | -1.0546 | 0.0240 | 1.6086, -1.6086 | 0.0246, 0.0246 |
| 34 | -0.4158 | 0.0237 | N/A | N/A |
| 35 | -0.2141 | 0.0188 | 0.7870, -0.7870 | 0.0208, 0.0208 |
| 36 | 0.3203 | 0.0194 | 0.6279, -0.6279 | 0.0226, 0.0226 |

Table 8.B.1 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b*-parameter Standard Error** | ***d*-parameters** | ***d*-parameters Standard Error** |
| 37 | -0.8136 | 0.0233 | 1.6458, -1.6458 | 0.0232, 0.0232 |
| 38 | -0.9154 | 0.0224 | 1.4140, -1.4140 | 0.0229, 0.0229 |
| 39 | 0.8797 | 0.0222 | -0.1380, 0.1380 | 0.0298, 0.0298 |
| 40 | -0.4413 | 0.0240 | N/A | N/A |
| 41 | -0.1531 | 0.0245 | N/A | N/A |
| 42 | -0.0224 | 0.0245 | N/A | N/A |
| 43 | 0.4379 | 0.0255 | N/A | N/A |
| 44 | 0.5496 | 0.0262 | N/A | N/A |
| 45 | -0.1067 | 0.0205 | 1.2645, -1.2645 | 0.0209, 0.0209 |
| 46 | 0.3956 | 0.0261 | N/A | N/A |
| 47 | 0.3262 | 0.0223 | 1.3383, -1.3383 | 0.0225, 0.0225 |
| 48 | 1.0445 | 0.0287 | N/A | N/A |
| 49 | 0.1685 | 0.0246 | N/A | N/A |
| 50 | -0.3923 | 0.0240 | N/A | N/A |
| 51 | 0.1881 | 0.0269 | N/A | N/A |
| 52 | 0.3632 | 0.0259 | N/A | N/A |

Table 8.B.2 IRT Item Difficulty for Grade Four

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b*-parameter Standard Error** | ***d*-parameters** | ***d*-parameters Standard Error** |
| 1 | 0.4270 | 0.0251 | N/A | N/A |
| 2 | -0.5392 | 0.0258 | N/A | N/A |
| 3 | -1.1145 | 0.0265 | N/A | N/A |
| 4 | -0.4356 | 0.0241 | 1.8786, -1.8786 | 0.0245, 0.0245 |
| 5 | 0.1737 | 0.0253 | N/A | N/A |
| 6 | 1.3498 | 0.0296 | N/A | N/A |
| 7 | 0.3105 | 0.0250 | N/A | N/A |
| 8 | 0.8286 | 0.0269 | N/A | N/A |
| 9 | 0.4431 | 0.0255 | N/A | N/A |
| 10 | -0.2731 | 0.0206 | 1.2319, -1.2319 | 0.0216, 0.0216 |
| 11 | -0.6588 | 0.0248 | N/A | N/A |
| 12 | -0.4054 | 0.0244 | N/A | N/A |
| 13 | 0.0309 | 0.0205 | 1.1450, -1.1450 | 0.0217, 0.0217 |
| 14 | -1.1525 | 0.0273 | N/A | N/A |
| 15 | 0.5191 | 0.0262 | N/A | N/A |
| 16 | -0.3305 | 0.0221 | 1.5977, -1.5977 | 0.0226, 0.0226 |
| 17 | 0.5164 | 0.0255 | N/A | N/A |
| 18 | -0.7325 | 0.0211 | 1.1629, -1.1629 | 0.0227, 0.0227 |
| 19 | 1.0697 | 0.0282 | N/A | N/A |
| 20 | 0.8251 | 0.0269 | N/A | N/A |
| 21 | -0.4195 | 0.0230 | 1.6843, -1.6843 | 0.0231, 0.0231 |
| 22 | 0.1623 | 0.0247 | N/A | N/A |
| 23 | 0.7236 | 0.0262 | N/A | N/A |
| 24 | 0.0730 | 0.0262 | N/A | N/A |
| 25 | -0.0384 | 0.0267 | N/A | N/A |
| 26 | 0.9309 | 0.0273 | N/A | N/A |
| 27 | 1.2228 | 0.0284 | N/A | N/A |
| 28 | -0.5419 | 0.0193 | 0.7374, -0.7374 | 0.0217, 0.0217 |
| 29 | -0.0592 | 0.0253 | N/A | N/A |
| 30 | 0.7670 | 0.0265 | N/A | N/A |
| 31 | -1.4180 | 0.0259 | 1.0804, -1.0804 | 0.0281, 0.0281 |
| 32 | -1.6339 | 0.0308 | 1.4146, -1.4146 | 0.0324, 0.0324 |
| 33 | 0.2138 | 0.0256 | N/A | N/A |
| 34 | -1.1563 | 0.0270 | N/A | N/A |
| 35 | -0.0727 | 0.0243 | N/A | N/A |
| 36 | 0.4908 | 0.0211 | 0.2685, -0.2685 | 0.0244, 0.0244 |
| 37 | 0.6680 | 0.0280 | N/A | N/A |
| 38 | 0.1316 | 0.0263 | N/A | N/A |

Table 8.B.2 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b*-parameter Standard Error** | ***d*-parameters** | ***d*-parameters Standard Error** |
| 39 | -0.4890 | 0.0196 | 0.8451, -0.8451 | 0.0215, 0.0215 |
| 40 | 0.5057 | 0.0259 | N/A | N/A |
| 41 | 0.3906 | 0.0255 | N/A | N/A |
| 42 | 0.3076 | 0.0232 | 1.4948, -1.4948 | 0.0235, 0.0235 |
| 43 | 0.3725 | 0.0209 | 1.0479, -1.0479 | 0.0224, 0.0224 |
| 44 | -0.0485 | 0.0246 | N/A | N/A |
| 45 | 0.2850 | 0.0250 | N/A | N/A |
| 46 | 0.2013 | 0.0250 | N/A | N/A |
| 47 | -1.0248 | 0.0212 | 0.7956, -0.7956 | 0.0242, 0.0242 |
| 48 | 0.1972 | 0.0250 | N/A | N/A |
| 49 | -0.3945 | 0.0246 | N/A | N/A |
| 50 | -0.1868 | 0.0248 | N/A | N/A |
| 51 | -0.0705 | 0.0245 | N/A | N/A |
| 52 | 0.3117 | 0.0252 | N/A | N/A |

Table 8.B.3 IRT Item Difficulty for Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b*-parameter Standard Error** | ***d*-parameters** | ***d-*parameters Standard Error** |
| 1 | 0.0891 | 0.0209 | 0.9947, -0.9947 | 0.0234, 0.0234 |
| 2 | -0.6675 | 0.0267 | N/A | N/A |
| 3 | -1.1062 | 0.0292 | N/A | N/A |
| 4 | -0.8237 | 0.0285 | N/A | N/A |
| 5 | -1.7795 | 0.0325 | N/A | N/A |
| 6 | -0.9722 | 0.0290 | N/A | N/A |
| 7 | -1.3228 | 0.0325 | 1.7585, -1.7585 | 0.0337, 0.0337 |
| 8 | 0.5861 | 0.0276 | N/A | N/A |
| 9 | -0.2998 | 0.0265 | N/A | N/A |
| 10 | 0.8714 | 0.0290 | N/A | N/A |
| 11 | -0.2540 | 0.0224 | 1.3666, -1.3666 | 0.0240, 0.0240 |
| 12 | 0.1900 | 0.0268 | N/A | N/A |
| 13 | 0.1877 | 0.0266 | N/A | N/A |
| 14 | 1.4279 | 0.0322 | N/A | N/A |
| 15 | 0.6887 | 0.0276 | N/A | N/A |
| 16 | 0.4296 | 0.0270 | N/A | N/A |
| 17 | 1.3152 | 0.0307 | N/A | N/A |
| 18 | 0.1017 | 0.0223 | 1.2850, -1.2850 | 0.0236, 0.0236 |
| 19 | 0.2516 | 0.0221 | 1.1850, -1.1850 | 0.0240, 0.0240 |
| 20 | 0.4967 | 0.0272 | N/A | N/A |
| 21 | 0.7403 | 0.0279 | N/A | N/A |
| 22 | 0.4484 | 0.0271 | N/A | N/A |
| 23 | 0.0086 | 0.0262 | N/A | N/A |
| 24 | -0.2043 | 0.0261 | N/A | N/A |
| 25 | 0.6131 | 0.0239 | 1.0018, -1.0018 | 0.0260, 0.0260 |
| 26 | 0.5181 | 0.0273 | N/A | N/A |
| 27 | -0.0615 | 0.0263 | N/A | N/A |
| 28 | 0.0372 | 0.0270 | N/A | N/A |
| 29 | 0.8161 | 0.0283 | N/A | N/A |
| 30 | 0.2614 | 0.0210 | 0.8011, -0.8011 | 0.0237, 0.0237 |
| 31 | -0.6227 | 0.0273 | N/A | N/A |
| 32 | -0.5875 | 0.0236 | 1.3845, -1.3845 | 0.0251, 0.0251 |
| 33 | -0.8745 | 0.0299 | N/A | N/A |
| 34 | -0.6388 | 0.0287 | N/A | N/A |
| 35 | 0.5557 | 0.0274 | N/A | N/A |
| 36 | 0.5470 | 0.0273 | N/A | N/A |
| 37 | 0.3722 | 0.0270 | N/A | N/A |
| 38 | 0.2362 | 0.0268 | N/A | N/A |

Table 8.B.3 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b*-parameter Standard Error** | ***d*-parameters** | ***d-*parameters Standard Error** |
| 39 | -0.0822 | 0.0262 | N/A | N/A |
| 40 | 0.0387 | 0.0267 | N/A | N/A |
| 41 | 0.9092 | 0.0289 | N/A | N/A |
| 42 | 1.0551 | 0.0294 | N/A | N/A |
| 43 | -0.2763 | 0.0265 | N/A | N/A |
| 44 | 0.5403 | 0.0221 | 0.6407, -0.6407 | 0.0254, 0.0254 |
| 45 | -0.7298 | 0.0255 | 1.6333, -1.6333 | 0.0266, 0.0266 |
| 46 | 0.3919 | 0.0278 | N/A | N/A |
| 47 | 0.1566 | 0.0266 | N/A | N/A |
| 48 | -0.0930 | 0.0269 | N/A | N/A |
| 49 | 0.0152 | 0.0269 | N/A | N/A |
| 50 | 0.1880 | 0.0201 | 0.6446, -0.6446 | 0.0236, 0.0236 |
| 51 | 1.4197 | 0.0335 | N/A | N/A |
| 52 | -1.4771 | 0.0269 | 0.5605, -0.5605 | 0.0319, 0.0319 |

Table 8.B.4 IRT Item Difficulty for Grade Six

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b*-parameter Standard Error** | ***d-*parameters** | ***d*-parameters Standard Error** |
| 1 | 0.2585 | 0.0321 | N/A | N/A |
| 2 | -0.5002 | 0.0327 | N/A | N/A |
| 3 | -0.1988 | 0.0256 | 0.8070, -0.8070 | 0.0288, 0.0288 |
| 4 | 0.8364 | 0.0343 | N/A | N/A |
| 5 | -0.4412 | 0.0330 | N/A | N/A |
| 6 | 0.4022 | 0.0328 | N/A | N/A |
| 7 | -0.7358 | 0.0337 | N/A | N/A |
| 8 | -1.5435 | 0.0412 | 1.4382, -1.4382 | 0.0433, 0.0433 |
| 9 | -0.4378 | 0.0323 | N/A | N/A |
| 10 | 1.4251 | 0.0379 | N/A | N/A |
| 11 | 0.3243 | 0.0251 | 0.7076, -0.7076 | 0.0284, 0.0284 |
| 12 | 0.0399 | 0.0323 | N/A | N/A |
| 13 | -0.8229 | 0.0290 | 1.1137, -1.1137 | 0.0313, 0.0313 |
| 14 | -1.0912 | 0.0348 | N/A | N/A |
| 15 | -0.1216 | 0.0320 | N/A | N/A |
| 16 | -1.4618 | 0.0329 | 0.3636, -0.3636 | 0.0398, 0.0398 |
| 17 | 0.4950 | 0.0337 | N/A | N/A |
| 18 | 0.5503 | 0.0334 | N/A | N/A |
| 19 | 0.0415 | 0.0331 | N/A | N/A |
| 20 | -1.0304 | 0.0279 | 0.5419, -0.5419 | 0.0329, 0.0329 |
| 21 | 0.1529 | 0.0326 | N/A | N/A |
| 22 | 0.2509 | 0.0329 | N/A | N/A |
| 23 | 0.2005 | 0.0323 | N/A | N/A |
| 24 | 1.0983 | 0.0352 | N/A | N/A |
| 25 | -1.4523 | 0.0377 | N/A | N/A |
| 26 | 0.2812 | 0.0327 | N/A | N/A |
| 27 | 0.5796 | 0.0331 | N/A | N/A |
| 28 | -0.8896 | 0.0338 | N/A | N/A |
| 29 | 0.0216 | 0.0324 | N/A | N/A |
| 30 | -0.7556 | 0.0292 | 1.2115, -1.2115 | 0.0312, 0.0312 |
| 31 | 1.1279 | 0.0365 | N/A | N/A |
| 32 | -0.1348 | 0.0330 | N/A | N/A |
| 33 | -0.1017 | 0.0325 | N/A | N/A |
| 34 | 0.3611 | 0.0326 | N/A | N/A |
| 35 | -1.1294 | 0.0346 | 1.4110, -1.4110 | 0.0371, 0.0371 |
| 36 | -0.5178 | 0.0263 | 0.8731, -0.8731 | 0.0289, 0.0289 |
| 37 | -0.5737 | 0.0344 | N/A | N/A |
| 38 | -0.6439 | 0.0341 | N/A | N/A |

Table 8.B.4 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b*-parameter Standard Error** | ***d-*parameters** | ***d*-parameters Standard Error** |
| 39 | -0.0933 | 0.0257 | 0.9694, -0.9694 | 0.0276, 0.0276 |
| 40 | 0.9002 | 0.0354 | N/A | N/A |
| 41 | -0.2478 | 0.0322 | N/A | N/A |
| 42 | -0.0107 | 0.0319 | N/A | N/A |
| 43 | 0.0407 | 0.0321 | N/A | N/A |
| 44 | -0.3482 | 0.0342 | N/A | N/A |
| 45 | 0.2863 | 0.0324 | N/A | N/A |
| 46 | 0.8924 | 0.0348 | N/A | N/A |
| 47 | 0.3781 | 0.0259 | 0.9128, -0.9128 | 0.0283, 0.0283 |
| 48 | 0.1679 | 0.0274 | 1.2252, -1.2252 | 0.0283, 0.0283 |
| 49 | -0.1995 | 0.0323 | N/A | N/A |
| 50 | 0.5984 | 0.0331 | N/A | N/A |
| 51 | 0.2847 | 0.0257 | 0.8680, -0.8680 | 0.0281, 0.0281 |
| 52 | 0.4801 | 0.0328 | N/A | N/A |

Table 8.B.5 IRT Item Difficulty for Grade Seven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b*-parameter Standard Error** | ***d*-parameters** | ***d*-parameters Standard Error** |
| 1 | -0.5048 | 0.0376 | N/A | N/A |
| 2 | 0.4787 | 0.0290 | 0.5358, -0.5358 | 0.0357, 0.0357 |
| 3 | -0.3845 | 0.0322 | 1.2834, -1.2834 | 0.0352, 0.0352 |
| 4 | -1.3298 | 0.0447 | N/A | N/A |
| 5 | -0.2113 | 0.0379 | N/A | N/A |
| 6 | 0.5877 | 0.0397 | N/A | N/A |
| 7 | -0.7147 | 0.0391 | N/A | N/A |
| 8 | 0.4657 | 0.0385 | N/A | N/A |
| 9 | 0.1432 | 0.0388 | N/A | N/A |
| 10 | 0.5056 | 0.0387 | N/A | N/A |
| 11 | -0.0655 | 0.0307 | 1.1750, -1.1750 | 0.0340, 0.0340 |
| 12 | 0.3262 | 0.0379 | N/A | N/A |
| 13 | 0.8622 | 0.0403 | N/A | N/A |
| 14 | 0.6494 | 0.0390 | N/A | N/A |
| 15 | 0.0841 | 0.0374 | N/A | N/A |
| 16 | -0.2897 | 0.0317 | 1.2506, -1.2506 | 0.0347, 0.0347 |
| 17 | -0.0336 | 0.0378 | N/A | N/A |
| 18 | -0.0724 | 0.0394 | N/A | N/A |
| 19 | 0.5648 | 0.0390 | N/A | N/A |
| 20 | -0.1653 | 0.0378 | N/A | N/A |
| 21 | 0.8046 | 0.0402 | N/A | N/A |
| 22 | 0.6494 | 0.0401 | N/A | N/A |
| 23 | -0.2113 | 0.0375 | N/A | N/A |
| 24 | 0.2264 | 0.0380 | N/A | N/A |
| 25 | -0.3096 | 0.0375 | N/A | N/A |
| 26 | -0.3478 | 0.0382 | N/A | N/A |
| 27 | -1.0483 | 0.0417 | 1.6194, -1.6194 | 0.0440, 0.0440 |
| 28 | -0.9749 | 0.0367 | 1.2259, -1.2259 | 0.0405, 0.0405 |
| 29 | 0.5941 | 0.0390 | N/A | N/A |
| 30 | 0.0417 | 0.0373 | N/A | N/A |
| 31 | 0.3141 | 0.0379 | N/A | N/A |
| 32 | -0.8917 | 0.0337 | 0.9550, -0.9550 | 0.0387, 0.0387 |
| 33 | 0.3613 | 0.0385 | N/A | N/A |
| 34 | -0.7688 | 0.0315 | 0.8545, -0.8545 | 0.0370, 0.0370 |
| 35 | -0.7193 | 0.0367 | 1.5551, -1.5551 | 0.0394, 0.0394 |
| 36 | 0.6728 | 0.0396 | N/A | N/A |
| 37 | 1.2621 | 0.0456 | N/A | N/A |
| 38 | 0.3165 | 0.0384 | N/A | N/A |

Table 8.B.5 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b*-parameter Standard Error** | ***d*-parameters** | ***d*-parameters Standard Error** |
| 39 | 0.7029 | 0.0345 | 1.2002, -1.2002 | 0.0377, 0.0377 |
| 40 | 0.5043 | 0.0388 | N/A | N/A |
| 41 | -0.2219 | 0.0374 | N/A | N/A |
| 42 | 0.6066 | 0.0390 | N/A | N/A |
| 43 | 3.2438 | 0.0850 | N/A | N/A |
| 44 | 1.1224 | 0.0438 | N/A | N/A |
| 45 | -0.4343 | 0.0272 | 0.0872, -0.0872 | 0.0377, 0.0377 |
| 46 | 0.1179 | 0.0390 | N/A | N/A |
| 47 | -0.0623 | 0.0375 | N/A | N/A |
| 48 | 1.1071 | 0.0425 | N/A | N/A |
| 49 | 0.1602 | 0.0306 | 1.0651, -1.0651 | 0.0340, 0.0340 |
| 50 | -0.1094 | 0.0295 | 0.8702, -0.8702 | 0.0333, 0.0333 |
| 51 | 0.1831 | 0.0379 | N/A | N/A |
| 52 | 0.5178 | 0.0391 | N/A | N/A |

Table 8.B.6 IRT Item Difficulty for Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b-*parameter Standard Error** | ***d*-parameters** | ***d*-parameters Standard Error** |
| 1 | -1.4332 | 0.0487 | N/A | N/A |
| 2 | -0.6906 | 0.0348 | 0.7589, -0.7589 | 0.0396, 0.0396 |
| 3 | -0.0524 | 0.0417 | N/A | N/A |
| 4 | -1.5642 | 0.0512 | N/A | N/A |
| 5 | -1.2151 | 0.0507 | 1.7884, -1.7884 | 0.0527, 0.0527 |
| 6 | -0.4409 | 0.0423 | N/A | N/A |
| 7 | -0.0453 | 0.0428 | N/A | N/A |
| 8 | 0.0567 | 0.0428 | N/A | N/A |
| 9 | 0.0170 | 0.0421 | N/A | N/A |
| 10 | -0.6606 | 0.0350 | 0.8802, -0.8802 | 0.0394, 0.0394 |
| 11 | 0.1350 | 0.0421 | N/A | N/A |
| 12 | 0.2640 | 0.0417 | N/A | N/A |
| 13 | -0.6587 | 0.0353 | 0.8887, -0.8887 | 0.0391, 0.0391 |
| 14 | 0.4455 | 0.0426 | N/A | N/A |
| 15 | -0.1076 | 0.0417 | N/A | N/A |
| 16 | 0.9460 | 0.0452 | N/A | N/A |
| 17 | 0.3864 | 0.0430 | N/A | N/A |
| 18 | -0.5836 | 0.0442 | 1.9877, -1.9877 | 0.0446, 0.0446 |
| 19 | 0.1683 | 0.0322 | 0.6751, -0.6751 | 0.0369, 0.0369 |
| 20 | 0.8010 | 0.0448 | N/A | N/A |
| 21 | -0.6839 | 0.0436 | N/A | N/A |
| 22 | -0.3955 | 0.0356 | 1.2577, -1.2577 | 0.0381, 0.0381 |
| 23 | -0.6459 | 0.0426 | N/A | N/A |
| 24 | -0.4386 | 0.0422 | N/A | N/A |
| 25 | 0.7559 | 0.0455 | N/A | N/A |
| 26 | 0.5047 | 0.0428 | N/A | N/A |
| 27 | 0.4449 | 0.0429 | N/A | N/A |
| 28 | -0.7428 | 0.0372 | 1.1463, -1.1463 | 0.0400, 0.0400 |
| 29 | 0.0293 | 0.0425 | N/A | N/A |
| 30 | 0.7376 | 0.0452 | N/A | N/A |
| 31 | 0.5251 | 0.0432 | N/A | N/A |
| 32 | 0.4935 | 0.0435 | N/A | N/A |
| 33 | -0.7170 | 0.0449 | N/A | N/A |
| 34 | 1.0515 | 0.0466 | N/A | N/A |
| 35 | -0.4740 | 0.0347 | 1.0203, -1.0203 | 0.0387, 0.0387 |
| 36 | 0.3381 | 0.0446 | N/A | N/A |
| 37 | 0.1987 | 0.0432 | N/A | N/A |
| 38 | 0.0686 | 0.0423 | N/A | N/A |

Table 8.B.6 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b-*parameter Standard Error** | ***d*-parameters** | ***d*-parameters Standard Error** |
| 39 | -0.0620 | 0.0447 | N/A | N/A |
| 40 | 0.5202 | 0.0437 | N/A | N/A |
| 41 | -0.3992 | 0.0437 | N/A | N/A |
| 42 | -0.1344 | 0.0425 | N/A | N/A |
| 43 | 0.3203 | 0.0381 | 1.4793, -1.4793 | 0.0392, 0.0392 |
| 44 | -0.0655 | 0.0329 | 0.9227, -0.9227 | 0.0364, 0.0364 |
| 45 | 1.1696 | 0.0483 | N/A | N/A |
| 46 | 0.3464 | 0.0451 | N/A | N/A |
| 47 | -0.4202 | 0.0416 | 1.8666, -1.8666 | 0.0424, 0.0424 |
| 48 | 0.9880 | 0.0480 | N/A | N/A |
| 49 | 0.2073 | 0.0429 | N/A | N/A |
| 50 | 0.3552 | 0.0431 | N/A | N/A |
| 51 | 0.2751 | 0.0423 | N/A | N/A |
| 52 | 0.6164 | 0.0432 | N/A | N/A |

Table 8.B.7 IRT Item Difficulty for High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b-*parameter Standard Error** | ***d*-parameters** | ***d*-parameters Standard Error** |
| 1 | 0.9623 | 0.0433 | N/A | N/A |
| 2 | -1.3312 | 0.0478 | 1.6297, -1.6297 | 0.0500, 0.0500 |
| 3 | -0.3532 | 0.0397 | N/A | N/A |
| 4 | -0.1169 | 0.0386 | N/A | N/A |
| 5 | 0.1782 | 0.0390 | N/A | N/A |
| 6 | -0.0132 | 0.0396 | N/A | N/A |
| 7 | -0.8880 | 0.0380 | 1.4343, -1.4343 | 0.0412, 0.0412 |
| 8 | -0.5451 | 0.0338 | 1.2806, -1.2806 | 0.0369, 0.0369 |
| 9 | -0.1382 | 0.0304 | 0.8661, -0.8661 | 0.0346, 0.0346 |
| 10 | 0.3197 | 0.0399 | N/A | N/A |
| 11 | -0.0126 | 0.0400 | N/A | N/A |
| 12 | 1.6840 | 0.0506 | N/A | N/A |
| 13 | 1.2866 | 0.0457 | N/A | N/A |
| 14 | -0.6966 | 0.0406 | N/A | N/A |
| 15 | -0.3799 | 0.0392 | N/A | N/A |
| 16 | -0.5526 | 0.0393 | N/A | N/A |
| 17 | 0.3826 | 0.0399 | N/A | N/A |
| 18 | -0.1536 | 0.0390 | N/A | N/A |
| 19 | -0.2150 | 0.0392 | N/A | N/A |
| 20 | -0.3345 | 0.0303 | 0.8030, -0.8030 | 0.0355, 0.0355 |
| 21 | -0.0360 | 0.0394 | N/A | N/A |
| 22 | 0.1142 | 0.0398 | N/A | N/A |
| 23 | 0.1339 | 0.0395 | N/A | N/A |
| 24 | 0.8456 | 0.0421 | N/A | N/A |
| 25 | 0.7415 | 0.0420 | N/A | N/A |
| 26 | 0.5772 | 0.0409 | N/A | N/A |
| 27 | 1.1902 | 0.0470 | N/A | N/A |
| 28 | 0.2026 | 0.0398 | N/A | N/A |
| 29 | -0.5408 | 0.0328 | 1.0353, -1.0353 | 0.0363, 0.0363 |
| 30 | 0.1818 | 0.0393 | N/A | N/A |
| 31 | 0.0655 | 0.0389 | N/A | N/A |
| 32 | -0.5766 | 0.0401 | N/A | N/A |
| 33 | 0.5055 | 0.0423 | N/A | N/A |
| 34 | -0.1717 | 0.0296 | 0.6912, -0.6912 | 0.0348, 0.0348 |
| 35 | 0.7832 | 0.0420 | N/A | N/A |
| 36 | 1.0822 | 0.0438 | N/A | N/A |
| 37 | 1.0723 | 0.0438 | N/A | N/A |
| 38 | -0.9124 | 0.0324 | 0.5301, -0.5301 | 0.0402, 0.0402 |

Table 8.B.7 *(continuation)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item Sequence** | ***b*-parameter** | ***b-*parameter Standard Error** | ***d*-parameters** | ***d*-parameters Standard Error** |
| 39 | 0.1975 | 0.0395 | N/A | N/A |
| 40 | 0.7581 | 0.0415 | N/A | N/A |
| 41 | 0.3352 | 0.0362 | 1.4947, -1.4947 | 0.0377, 0.0377 |
| 42 | 0.6293 | 0.0357 | 1.2147, -1.2147 | 0.0386, 0.0386 |
| 43 | 1.1743 | 0.0446 | N/A | N/A |
| 44 | 0.2948 | 0.0398 | N/A | N/A |
| 45 | 0.3254 | 0.0408 | N/A | N/A |
| 46 | 1.1042 | 0.0456 | N/A | N/A |
| 47 | 0.1519 | 0.0404 | N/A | N/A |
| 48 | 0.5219 | 0.0408 | N/A | N/A |
| 49 | 0.0909 | 0.0394 | N/A | N/A |
| 50 | 0.0083 | 0.0399 | N/A | N/A |
| 51 | 0.1298 | 0.0397 | N/A | N/A |
| 52 | 0.7646 | 0.0421 | N/A | N/A |

### Appendix 8.C: Testing Time Analyses

**Notes:**

* Total raw scores were used to partition students into quartiles.
* All students who completed testing with a valid reporting scale score are included.
* Form 1 represents the general form, while Form A represents the accommodated form.

Table 8.C.1 Total Testing Time (in Minutes) at Each Raw Score Interval—Grade Three

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Form** | **Raw Score Interval** | **N** | **Mean** | **SD** | **Minimum** | **Maximum** | **% Pt. 1** | **% Pt. 10** | **% Pt. 25** | **% Pt. 50** | **% Pt. 75** | **% Pt. 90** | **% Pt. 99** |
| 1 | 0–22 | 2,441 | 74.30 | 57.16 | 9.14 | 609.01 | 12.96 | 24.76 | 38.17 | 60.68 | 92.07 | 138.36 | 288.84 |
| 1 | 23–26 | 2,090 | 83.20 | 64.78 | 4.65 | 671.49 | 13.01 | 27.76 | 42.90 | 66.80 | 104.40 | 152.26 | 333.80 |
| 1 | 27–35 | 2,693 | 104.24 | 71.16 | 9.57 | 726.68 | 20.09 | 42.10 | 59.90 | 88.30 | 124.68 | 180.74 | 377.40 |
| 1 | 36–66 | 2,604 | 116.53 | 68.72 | 18.38 | 770.34 | 37.40 | 58.07 | 74.82 | 99.61 | 135.47 | 190.10 | 402.08 |
| A | 0–16 | 306 | 61.70 | 37.79 | 6.11 | 208.65 | 12.99 | 22.04 | 33.95 | 53.61 | 81.31 | 111.08 | 193.90 |
| A | 17–19 | 267 | 69.23 | 46.00 | 8.67 | 310.13 | 12.03 | 25.42 | 38.33 | 59.25 | 87.11 | 118.18 | 265.20 |
| A | 20–25 | 362 | 80.83 | 52.07 | 7.09 | 329.48 | 13.42 | 31.13 | 46.88 | 68.83 | 98.00 | 138.29 | 300.36 |
| A | 26–60 | 342 | 90.76 | 42.53 | 22.77 | 277.46 | 33.62 | 48.19 | 60.40 | 81.95 | 109.50 | 138.91 | 260.69 |

Table 8.C.2 Total Testing Time (in Minutes) at Each Raw Score Interval—Grade Four

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Form** | **Raw Score Interval** | **N** | **Mean** | **SD** | **Minimum** | **Maximum** | **% Pt. 1** | **% Pt. 10** | **% Pt. 25** | **% Pt. 50** | **% Pt. 75** | **% Pt. 90** | **% Pt. 99** |
| 1 | 0–23 | 2,074 | 82.36 | 53.57 | 5.40 | 441.01 | 13.15 | 28.70 | 46.05 | 69.93 | 104.54 | 150.27 | 281.29 |
| 1 | 24–28 | 2,218 | 94.63 | 64.89 | 8.19 | 1020.43 | 16.81 | 36.49 | 54.54 | 81.39 | 117.13 | 163.67 | 344.42 |
| 1 | 29–36 | 2,431 | 115.33 | 64.41 | 16.71 | 658.69 | 29.23 | 52.04 | 72.39 | 100.86 | 140.76 | 195.18 | 348.41 |
| 1 | 37–66 | 2,440 | 122.79 | 59.89 | 19.98 | 559.58 | 39.84 | 64.74 | 82.79 | 109.42 | 146.85 | 196.52 | 339.13 |
| A | 0–19 | 234 | 72.92 | 51.95 | 4.95 | 364.01 | 9.52 | 24.24 | 40.45 | 58.01 | 88.21 | 131.74 | 245.92 |
| A | 20–23 | 210 | 75.99 | 48.11 | 10.06 | 301.46 | 14.21 | 31.02 | 45.07 | 63.21 | 93.42 | 140.90 | 215.66 |
| A | 24–31 | 273 | 89.57 | 54.50 | 18.53 | 407.77 | 22.13 | 39.95 | 53.78 | 77.33 | 105.05 | 163.76 | 306.90 |
| A | 32–61 | 239 | 103.69 | 51.54 | 33.11 | 351.50 | 39.76 | 54.17 | 68.20 | 88.76 | 127.77 | 175.52 | 300.40 |

Table 8.C.3 Total Testing Time (in Minutes) at Each Raw Score Interval—Grade Five

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Form** | **Raw Score Interval** | **N** | **Mean** | **SD** | **Minimum** | **Maximum** | **% Pt. 1** | **% Pt. 10** | **% Pt. 25** | **% Pt. 50** | **% Pt. 75** | **% Pt. 90** | **% Pt. 99** |
| 1 | 0–22 | 1,891 | 83.93 | 54.80 | 5.05 | 582.53 | 12.32 | 31.04 | 45.85 | 72.37 | 107.64 | 148.83 | 280.85 |
| 1 | 23–27 | 1,728 | 96.87 | 62.80 | 8.66 | 766.76 | 18.74 | 38.70 | 56.75 | 83.01 | 121.59 | 170.31 | 305.40 |
| 1 | 28–35 | 2,154 | 115.78 | 65.71 | 15.09 | 745.27 | 29.00 | 52.70 | 72.28 | 102.59 | 143.02 | 194.78 | 353.57 |
| 1 | 36–64 | 1,990 | 119.30 | 64.25 | 31.87 | 1000.89 | 41.94 | 62.30 | 78.72 | 106.14 | 144.19 | 188.35 | 294.92 |
| A | 0–20 | 156 | 66.51 | 40.76 | 4.36 | 270.69 | 12.32 | 23.83 | 37.97 | 57.39 | 86.86 | 112.56 | 242.65 |
| A | 21–24 | 153 | 70.60 | 34.21 | 5.60 | 223.93 | 17.27 | 35.85 | 44.61 | 67.04 | 89.53 | 113.84 | 180.40 |
| A | 25–30 | 180 | 76.27 | 31.08 | 13.31 | 206.85 | 18.85 | 41.07 | 51.50 | 71.55 | 98.25 | 118.93 | 160.82 |
| A | 31–63 | 182 | 85.87 | 33.85 | 28.20 | 235.51 | 28.48 | 49.12 | 63.76 | 79.60 | 101.36 | 129.63 | 185.02 |

Table 8.C.4 Total Testing Time (in Minutes) at Each Raw Score Interval—Grade Six

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Form** | **Raw Score Interval** | **N** | **Mean** | **SD** | **Minimum** | **Maximum** | **% Pt. 1** | **% Pt. 10** | **% Pt. 25** | **% Pt. 50** | **% Pt. 75** | **% Pt. 90** | **% Pt. 99** |
| 1 | 0–25 | 1,320 | 88.82 | 53.48 | 7.71 | 471.17 | 16.07 | 34.77 | 52.77 | 77.91 | 112.50 | 154.11 | 268.10 |
| 1 | 26–32 | 1,350 | 108.88 | 61.13 | 15.70 | 495.63 | 24.77 | 50.12 | 68.35 | 95.90 | 133.61 | 180.09 | 337.38 |
| 1 | 33–40 | 1,323 | 121.83 | 67.13 | 27.39 | 684.04 | 37.09 | 60.89 | 80.25 | 106.05 | 145.16 | 197.57 | 389.57 |
| 1 | 41–65 | 1,423 | 121.28 | 65.23 | 32.61 | 577.46 | 43.14 | 64.42 | 80.79 | 104.79 | 138.84 | 195.44 | 383.65 |
| A | 0–24 | 112 | 89.15 | 62.03 | 13.02 | 337.35 | 20.34 | 30.71 | 51.85 | 74.34 | 112.22 | 145.85 | 304.97 |
| A | 25–30 | 111 | 106.95 | 74.47 | 27.28 | 487.38 | 32.06 | 48.64 | 64.02 | 82.49 | 129.12 | 164.92 | 462.00 |
| A | 31–38 | 132 | 117.25 | 77.33 | 34.35 | 461.34 | 41.41 | 57.68 | 73.09 | 94.63 | 126.97 | 215.41 | 389.96 |
| A | 39–64 | 121 | 123.79 | 74.25 | 33.70 | 442.79 | 35.20 | 58.89 | 75.12 | 104.42 | 149.68 | 210.23 | 396.42 |

Table 8.C.5 Total Testing Time (in Minutes) at Each Raw Score Interval—Grade Seven

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Form** | **Raw Score Interval** | **N** | **Mean** | **SD** | **Minimum** | **Maximum** | **% Pt. 1** | **% Pt. 10** | **% Pt. 25** | **% Pt. 50** | **% Pt. 75** | **% Pt. 90** | **% Pt. 99** |
| 1 | 0–23 | 819 | 81.61 | 48.42 | 6.64 | 387.92 | 16.86 | 30.37 | 44.71 | 73.30 | 103.53 | 148.73 | 218.27 |
| 1 | 24–28 | 878 | 91.26 | 47.28 | 8.71 | 295.36 | 21.10 | 39.26 | 55.74 | 83.55 | 117.77 | 154.99 | 240.41 |
| 1 | 29–36 | 1,067 | 106.70 | 46.58 | 15.84 | 443.24 | 29.82 | 54.16 | 73.27 | 100.52 | 132.65 | 169.25 | 245.10 |
| 1 | 37–65 | 961 | 112.52 | 43.27 | 23.73 | 345.69 | 41.97 | 65.58 | 82.00 | 104.59 | 136.89 | 167.45 | 250.15 |
| A | 0–23 | 26 | 86.70 | 41.00 | 30.78 | 211.66 | 30.78 | 32.34 | 63.83 | 82.08 | 100.81 | 123.06 | 211.66 |
| A | 24–27 | 21 | 86.66 | 38.04 | 31.98 | 161.16 | 31.98 | 43.67 | 55.69 | 84.07 | 116.21 | 135.25 | 161.16 |
| A | 28–33 | 31 | 108.00 | 59.85 | 30.25 | 353.40 | 30.25 | 53.99 | 72.54 | 94.66 | 135.88 | 158.57 | 353.40 |
| A | 34–63 | 30 | 104.91 | 32.91 | 36.89 | 189.34 | 36.89 | 53.36 | 91.87 | 103.92 | 117.77 | 140.73 | 189.34 |

Table 8.C.6 Total Testing Time (in Minutes) at Each Raw Score Interval—Grade Eight

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Form** | **Raw Score Interval** | **N** | **Mean** | **SD** | **Minimum** | **Maximum** | **% Pt. 1** | **% Pt. 10** | **% Pt. 25** | **% Pt. 50** | **% Pt. 75** | **% Pt. 90** | **% Pt. 99** |
| 1 | 0–23 | 642 | 74.11 | 43.38 | 5.15 | 312.13 | 10.14 | 28.05 | 44.30 | 66.84 | 94.60 | 132.50 | 215.05 |
| 1 | 24–30 | 933 | 93.89 | 46.06 | 7.48 | 362.21 | 18.92 | 42.63 | 61.26 | 88.72 | 118.96 | 148.94 | 239.53 |
| 1 | 31–37 | 711 | 111.48 | 44.18 | 24.49 | 325.26 | 33.49 | 60.51 | 79.43 | 105.37 | 137.00 | 167.83 | 239.53 |
| 1 | 38–64 | 868 | 126.21 | 44.73 | 37.60 | 372.09 | 49.12 | 75.83 | 95.88 | 120.48 | 147.18 | 181.78 | 271.01 |
| A | 0–18 | 20 | 67.85 | 29.92 | 10.74 | 129.83 | 10.74 | 28.60 | 46.86 | 64.22 | 85.10 | 110.14 | 129.83 |
| A | 19–24 | 22 | 93.98 | 35.82 | 28.47 | 170.25 | 28.47 | 47.40 | 77.10 | 90.27 | 118.48 | 128.50 | 170.25 |
| A | 25–30 | 20 | 105.19 | 28.75 | 42.47 | 146.96 | 42.47 | 71.01 | 85.17 | 101.30 | 134.25 | 143.40 | 146.96 |
| A | 31–58 | 23 | 108.56 | 25.39 | 66.48 | 153.45 | 66.48 | 68.51 | 96.21 | 108.22 | 126.94 | 136.85 | 153.45 |

Table 8.C.7 Total Testing Time (in Minutes) at Each Raw Score Interval—High School, All Grade Levels

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Form** | **Raw Score Interval** | **N** | **Mean** | **SD** | **Minimum** | **Maximum** | **% Pt. 1** | **% Pt. 10** | **% Pt. 25** | **% Pt. 50** | **% Pt. 75** | **% Pt. 90** | **% Pt. 99** |
| 1 | 0–20 | 772 | 63.71 | 40.53 | 3.67 | 368.19 | 9.00 | 22.58 | 37.81 | 56.77 | 79.44 | 109.57 | 199.84 |
| 1 | 21–26 | 923 | 76.31 | 34.64 | 4.55 | 248.77 | 13.50 | 38.58 | 51.78 | 72.71 | 93.27 | 121.22 | 181.88 |
| 1 | 27–34 | 897 | 93.42 | 39.38 | 21.13 | 448.86 | 30.74 | 53.02 | 67.93 | 86.48 | 111.62 | 139.05 | 223.60 |
| 1 | 35–62 | 907 | 101.05 | 38.04 | 27.22 | 319.57 | 40.23 | 61.57 | 75.03 | 94.21 | 119.18 | 147.34 | 229.67 |
| A | 0–26 | 3 | 71.07 | 26.16 | 41.38 | 90.75 | 41.38 | 41.38 | 41.38 | 81.08 | 90.75 | 90.75 | 90.75 |
| A | 27–33 | 4 | 53.78 | 24.50 | 26.95 | 74.61 | 26.95 | 26.95 | 33.00 | 56.77 | 74.56 | 74.61 | 74.61 |
| A | 34–40 | 3 | 77.60 | 31.94 | 41.68 | 102.78 | 41.68 | 41.68 | 41.68 | 88.35 | 102.78 | 102.78 | 102.78 |
| A | 41–62 | 4 | 107.77 | 26.46 | 81.88 | 138.58 | 81.88 | 81.88 | 85.94 | 105.30 | 129.59 | 138.58 | 138.58 |

### Appendix 8.D: Reliability Analyses

**Notes:**

* The reliabilities are reported only for samples that comprise 11 or more examinees.
* In some cases in appendix 8.D, score reliabilities were not estimable and are presented in the tables as “N/A.”
* Results based on samples that contain 50 or fewer examinees should be interpreted with caution because of small sample sizes.

Table 8.D.1 Reliabilities and SEMs by Gender

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Male N** | **Male Reliability** | **Male SEM** | **Female N** | **Female Reliability** | **Female SEM** |
| Grade 3 | 4,821 | 0.86 | 3.65 | 4,944 | 0.87 | 3.67 |
| Grade 4 | 4,472 | 0.85 | 3.59 | 4,664 | 0.85 | 3.59 |
| Grade 5 | 3,758 | 0.84 | 3.56 | 3,975 | 0.85 | 3.56 |
| Grade 6 | 2,582 | 0.86 | 3.64 | 2,797 | 0.86 | 3.62 |
| Grade 7 | 1,755 | 0.80 | 3.66 | 1,936 | 0.83 | 3.62 |
| Grade 8 | 1,487 | 0.84 | 3.59 | 1,638 | 0.85 | 3.56 |
| High school | 1,677 | 0.83 | 3.53 | 1,772 | 0.84 | 3.56 |

Table 8.D.2 Reliabilities and SEMs by Economic Status

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Not Disadvantaged N** | **Not Disadvantaged Reliability** | **Not Disadvantaged SEM** | **Disadvantaged N** | **Disadvantaged Reliability** | **Disadvantaged SEM** |
| Grade 3 | 2,952 | 0.88 | 3.66 | 6,815 | 0.85 | 3.66 |
| Grade 4 | 2,884 | 0.86 | 3.60 | 6,252 | 0.84 | 3.59 |
| Grade 5 | 2,458 | 0.85 | 3.57 | 5,277 | 0.84 | 3.55 |
| Grade 6 | 1,939 | 0.87 | 3.60 | 3,441 | 0.85 | 3.64 |
| Grade 7 | 1,232 | 0.84 | 3.63 | 2,461 | 0.81 | 3.65 |
| Grade 8 | 955 | 0.87 | 3.54 | 2,170 | 0.84 | 3.59 |
| High school | 846 | 0.84 | 3.54 | 2,603 | 0.84 | 3.55 |

Table 8.D.3 Reliabilities and SEMs by Special Education Services

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **No Special Services N** | **No Special Services Reliability** | **No Special Services SEM** | **Special Services N** | **Special Services Reliability** | **Special Services SEM** |
| Grade 3 | 9,152 | 0.87 | 3.66 | 615 | 0.79 | 3.63 |
| Grade 4 | 8,585 | 0.85 | 3.59 | 551 | 0.79 | 3.57 |
| Grade 5 | 7,303 | 0.84 | 3.56 | 432 | 0.79 | 3.53 |
| Grade 6 | 5,105 | 0.86 | 3.63 | 275 | 0.74 | 3.62 |
| Grade 7 | 3,538 | 0.82 | 3.64 | 155 | 0.71 | 3.59 |
| Grade 8 | 2,990 | 0.85 | 3.58 | 135 | 0.75 | 3.55 |
| High school | 3,335 | 0.84 | 3.55 | 114 | 0.70 | 3.42 |

Table 8.D.4 Reliabilities and SEMs by Attendance in US Schools

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Less than 12 Months N** | **Less than 12 Months Reliability** | **Less than 12 Months SEM** | **12 Months or More N** | **12 Months or More Reliability** | **12 Months or More SEM** |
| Grade 3 | 374 | 0.85 | 3.65 | 9,393 | 0.87 | 3.66 |
| Grade 4 | 342 | 0.83 | 3.54 | 8,794 | 0.85 | 3.59 |
| Grade 5 | 314 | 0.85 | 3.52 | 7,421 | 0.85 | 3.56 |
| Grade 6 | 265 | 0.85 | 3.59 | 5,115 | 0.86 | 3.63 |
| Grade 7 | 259 | 0.79 | 3.58 | 3,434 | 0.83 | 3.64 |
| Grade 8 | 258 | 0.84 | 3.61 | 2,867 | 0.85 | 3.57 |
| High school | 775 | 0.83 | 3.50 | 2,674 | 0.84 | 3.55 |

Table 8.D.5 Reliabilities and SEMs by Spanish Instruction Status

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **Received Instruction in Spanish N** | **Received Instruction in Spanish Reliability** | **Received Instruction in Spanish SEM** | **Not Received Instruction in Spanish N** | **Not Received Instruction in Spanish Reliability** | **Not Received Instruction in Spanish SEM** |
| Grade 3 | 9,288 | 0.87 | 3.66 | 479 | 0.73 | 3.62 |
| Grade 4 | 8,804 | 0.85 | 3.59 | 332 | 0.81 | 3.58 |
| Grade 5 | 7,497 | 0.85 | 3.56 | 238 | 0.80 | 3.57 |
| Grade 6 | 5,225 | 0.86 | 3.63 | 155 | 0.82 | 3.63 |
| Grade 7 | 3,551 | 0.83 | 3.64 | 142 | 0.70 | 3.60 |
| Grade 8 | 2,967 | 0.85 | 3.57 | 158 | 0.82 | 3.62 |
| High school | 3,029 | 0.83 | 3.55 | 420 | 0.86 | 3.51 |

Table 8.D.6 Reliabilities and SEMs by English Proficiency

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **English Only N** | **English Only Reliability** | **English Only SEM** | **Initial Fluent English Proficient (IFEP) N** | **IFEP Reliability** | **IFEP SEM** |
| Grade 3 | 3,531 | 0.86 | 3.65 | 862 | 0.89 | 3.64 |
| Grade 4 | 3,137 | 0.85 | 3.60 | 490 | 0.88 | 3.55 |
| Grade 5 | 2,448 | 0.84 | 3.56 | 406 | 0.87 | 3.51 |
| Grade 6 | 1,685 | 0.86 | 3.63 | 298 | 0.88 | 3.48 |
| Grade 7 | 997 | 0.84 | 3.62 | 175 | 0.86 | 3.58 |
| Grade 8 | 754 | 0.86 | 3.56 | 111 | 0.88 | 3.49 |
| High school | 458 | 0.82 | 3.53 | 104 | 0.88 | 3.49 |

Table 8.D.7 Reliabilities and SEMs by English Proficiency (Continued)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade Level or Grade Band** | **EL N** | **EL Reliability** | **EL SEM** | **RFEP N** | **RFEP Reliability** | **RFEP SEM** |
| Grade 3 | 4,337 | 0.83 | 3.66 | 1,035 | 0.89 | 3.62 |
| Grade 4 | 3,879 | 0.80 | 3.59 | 1,625 | 0.86 | 3.56 |
| Grade 5 | 2,736 | 0.79 | 3.56 | 2,142 | 0.84 | 3.53 |
| Grade 6 | 1,676 | 0.79 | 3.65 | 1,714 | 0.84 | 3.60 |
| Grade 7 | 1,056 | 0.73 | 3.64 | 1,459 | 0.82 | 3.63 |
| Grade 8 | 839 | 0.77 | 3.59 | 1,418 | 0.84 | 3.56 |
| High school | 1,358 | 0.81 | 3.51 | 1,527 | 0.84 | 3.55 |

### Appendix 8.E: Analyses of Classification

Table 8.E.1 Reliability of Classification for Grade Three: Decision Accuracy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 300–348 | 0.68 | 0.04 | 0.00 | 0.71 |
| 349–359 | 0.04 | 0.13 | 0.01 | 0.19 |
| 360–399 | 0.00 | 0.02 | 0.07 | 0.10 |

Estimated Proportion Correctly Classified: Total = 0.88

Table 8.E.2 Reliability of Classification for Grade Three: Decision Consistency

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 300–348 | 0.65 | 0.06 | 0.00 | 0.71 |
| 349–359 | 0.05 | 0.11 | 0.02 | 0.19 |
| 360–399 | 0.00 | 0.03 | 0.07 | 0.10 |

Estimated Proportion Correctly Classified: Total = 0.83

Table 8.E.3 Reliability of Classification for Grade Four: Decision Accuracy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 400–448 | 0.61 | 0.04 | 0.00 | 0.65 |
| 449–459 | 0.06 | 0.18 | 0.01 | 0.25 |
| 460–499 | 0.00 | 0.03 | 0.07 | 0.10 |

Estimated Proportion Correctly Classified: Total = 0.86

Table 8.E.4 Reliability of Classification for Grade Four: Decision Consistency

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 400–448 | 0.58 | 0.07 | 0.00 | 0.65 |
| 449–459 | 0.07 | 0.16 | 0.03 | 0.25 |
| 460–499 | 0.00 | 0.03 | 0.07 | 0.10 |

Estimated Proportion Correctly Classified: Total = 0.80

Table 8.E.5 Reliability of Classification for Grade Five: Decision Accuracy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 500–545 | 0.45 | 0.06 | 0.00 | 0.51 |
| 546–559 | 0.07 | 0.30 | 0.02 | 0.38 |
| 560–599 | 0.00 | 0.03 | 0.08 | 0.11 |

Estimated Proportion Correctly Classified: Total = 0.83

Table 8.E.6 Reliability of Classification for Grade Five: Decision Consistency

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 500–545 | 0.42 | 0.09 | 0.00 | 0.51 |
| 546–559 | 0.09 | 0.26 | 0.04 | 0.38 |
| 560–599 | 0.00 | 0.03 | 0.08 | 0.11 |

Estimated Proportion Correctly Classified: Total = 0.76

Table 8.E.7 Reliability of Classification for Grade Six: Decision Accuracy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 600–647 | 0.39 | 0.06 | 0.00 | 0.46 |
| 648–659 | 0.05 | 0.28 | 0.02 | 0.35 |
| 660–699 | 0.00 | 0.04 | 0.14 | 0.19 |

Estimated Proportion Correctly Classified: Total = 0.81

Table 8.E.8 Reliability of Classification for Grade Six: Decision Consistency

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 600–647 | 0.37 | 0.09 | 0.00 | 0.46 |
| 648–659 | 0.08 | 0.23 | 0.04 | 0.35 |
| 660–699 | 0.00 | 0.05 | 0.13 | 0.19 |

Estimated Proportion Correctly Classified: Total = 0.74

Table 8.E.9 Reliability of Classification for Grade Seven: Decision Accuracy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 700–743 | 0.56 | 0.06 | 0.00 | 0.62 |
| 744–759 | 0.06 | 0.26 | 0.00 | 0.33 |
| 760–799 | 0.00 | 0.02 | 0.03 | 0.06 |

Estimated Proportion Correctly Classified: Total = 0.85

Table 8.E.10 Reliability of Classification for Grade Seven: Decision Consistency

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 700–743 | 0.52 | 0.09 | 0.00 | 0.62 |
| 744–759 | 0.08 | 0.23 | 0.02 | 0.33 |
| 760–799 | 0.00 | 0.02 | 0.03 | 0.06 |

Estimated Proportion Correctly Classified: Total = 0.78

Table 8.E.11 Reliability of Classification for Grade Eight: Decision Accuracy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 800–847 | 0.55 | 0.05 | 0.00 | 0.61 |
| 848–859 | 0.06 | 0.20 | 0.02 | 0.28 |
| 860–899 | 0.00 | 0.03 | 0.08 | 0.12 |

Estimated Proportion Correctly Classified: Total = 0.84

Table 8.E.12 Reliability of Classification for Grade Eight: Decision Consistency

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 800–847 | 0.53 | 0.08 | 0.00 | 0.61 |
| 848–859 | 0.08 | 0.17 | 0.03 | 0.28 |
| 860–899 | 0.00 | 0.03 | 0.08 | 0.12 |

Estimated Proportion Correctly Classified: Total = 0.77

Table 8.E.13 Reliability of Classification for High School: Decision Accuracy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 900–949 | 0.60 | 0.05 | 0.00 | 0.66 |
| 950–959 | 0.06 | 0.17 | 0.01 | 0.24 |
| 960–999 | 0.01 | 0.04 | 0.06 | 0.11 |

Estimated Proportion Correctly Classified: Total = 0.84

Table 8.E.14 Reliability of Classification for High School: Decision Consistency

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporting Score** | **Score Reporting Range 1** | **Score Reporting Range 2** | **Score Reporting Range 3** | **Category Total** |
| 900–949 | 0.57 | 0.08 | 0.00 | 0.66 |
| 950–959 | 0.07 | 0.14 | 0.03 | 0.24 |
| 960–999 | 0.01 | 0.04 | 0.06 | 0.11 |

Estimated Proportion Correctly Classified: Total = 0.77

## 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 Spanish Assessment (CSA). As part of this effort, ETS staff worked with its Office of Professional Standards Compliance, which publishes and maintains the *ETS Standards for Quality and Fairness* (ETS, 2014). These *Standards* support the goals of delivering technically sound, fair, and useful products and services; and assisting the public and auditors evaluating those products and services. Quality control procedures are outlined in this chapter.

### Quality Control of Item Development

ETS’ goal is to provide the best standards-based and innovative items for the CSA. Items developed for the CSA were subject to an extensive item review process. The item writers responsible for developing CSA items were trained in California Assessment of Student Performance and Progress (CAASPP) and ETS policies on quality control of item content, bias and sensitivity guidelines, as well as guidelines for accessibility, to ensure that the items allow the widest possible range of students to demonstrate their abilities.

Once a draft item was accepted for authoring—that is, once it was entered into ETS’ item bank and formatted for use in an assessment—ETS employed a series of internal reviews and an initial CDE review. These reviews used established criteria and specifications to judge the quality of an item’s content and ensured that each item measured what it was intended to measure. These reviews also examined the overall quality of the test items before presentation to the CDE and item reviewers. To finish the process, a group of California educators reviewed the items for accessibility, bias and sensitivity, and content, and made recommendations for item enhancement. The details on quality control of item development are described in section [*3.2 ETS Item Review Process*](#_ETS_Item_Review).

During administrations of the CSA, when sufficient student response data on each item became available, ETS Psychometric Analysis & Research (PAR) staff conducted item analyses and a key check to examine whether the items performed as expected. ETS psychometric staff conducted a thorough evaluation of all item statistics using the statistical criteria described in subsection [*8.2.7 Summary of Classical Item Analyses Flagging Criteria*](#_Summary_of_Classical_1) to flag items that were potentially problematic because of poor item performance, content issues, item bias, or accessibility challenges. Flagged items were then reviewed by ETS Assessment and Learning Technology Research & Development (ALTRD) staff to determine whether issues existed.

### Quality Control of Test Assembly and Delivery

The assembly of all test forms must conform to blueprints that represent a set of constraints and specifications. ETS conducted multiple levels of quality assurance (QA) checks on each assembled CSA 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 item answers in the key were correctly identified and documented in the scoring system;
* items were scored correctly in the item bank and incorrect responses were scored as incorrect;
* all items assessed the intended standard;
* all content in the item was correct with the exception of distractors, which are intended to be incorrect;
* all items met the statistical criteria, to the extent possible;
* distractors were plausible;
* multiple-choice item options were parallel in structure;
* language was grade-level appropriate;
* no more than three multiple-choice items in a row had the same key;
* all graphics were correct (copyright, spelling, relevance, etc.);
* there were no unintended mechanical errors in grammar, spelling, punctuation, and the like; and
* items adhered to the approved style guide.

Reviews were also conducted for functionality and sequencing during the user acceptance testing (UAT) process to ensure all items functioned as expected. Three cycles of UAT were conducted: the first by the test delivery system (TDS) vendor, the second by ETS, and the third by the CDE. CDE staff made a final quality check to ensure that all issues identified during UAT were resolved before the release of the operational assessment.

### Quality Control of Test Materials

Brief descriptions of the types of materials used for and during testing appear in the following subsections.

#### Collecting Test Materials for Computer-based Assessments

During the 2021–22 CSA administration, there were no test materials to be collected as a result of computer-based testing.

#### Processing Test Materials for Computer-based Assessments

Computer-based tests submitted by students were transmitted from Cambium Assessment, Inc. (CAI) to ETS each day. Each system checked for the completeness of the student record and stopped records that were identified as having an error. (For example, the system would identify a test part that was missing a content registration ID, a unique identifier that matches the student’s opportunities.)

### Quality Control of Test Administration

The quality of test administration for the CSA was monitored and controlled through several strategies.

A fully supported Outreach team that includes California Technical Assistance Center phone support and Success Agents, supported all local educational agencies (LEAs) in the administration of the CAASPP. In addition to providing guidance and answering questions, the Outreach team regularly conducted campaigns on particular administration topics to ensure all LEAs understood correct test administration procedures. Outreach was guided by individuals who managed communications to LEAs; provided regional and web-based trainings; and hosted a website, [the](https://www.caaspp.org/) CAASPP website, that housed a full range of manuals, videos, and other instructional and support materials.

The quality of test administration was further managed through comprehensive rules and guidelines for maintaining the security and standardization of the CAASPP. LEAs received training on these topics and were provided tools for reporting security incidents and resolving testing discrepancies for specific testing sessions.

The ETS Office of Testing Integrity (OTI) reinforced the quality control procedures for test administration, providing quality assurance services for all testing programs managed by ETS. The detailed procedures the OTI developed and applied in quality control are described in subsection [*5.6.1 ETS’ Office of Testing Integrity*](#_ETS’_Office_of_1).

### Quality Control of Scoring

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

#### Machine-Scoring Procedures

To ensure valid item-level scoring for the CSA, quality control procedures were employed by CAI, the CAASPP subcontractor responsible for providing the TDS and scoring machine-scorable items. CAI staff independently reviewed all CSA forms by producing sample results for tests. The sample results were compared with the answer keys for each form to confirm the accuracy of scoring keys. The scores for all applicable items were recorded. A final comparison of the test map to each computer-based form as configured in the UAT environment ensured that no changes to the form were introduced prior to operational deployment.

A real-time, quality-monitoring component was built into the TDS. After 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.

#### Development of Scoring Specifications

A number of measures were taken to ascertain that the scoring keys were applied to the student responses as intended and the student scores were computed accurately. ETS built and reviewed the scoring system models based on the reporting specifications approved by the CDE. These specifications contain detailed scoring procedures, along with the procedures for determining whether a student has attempted 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 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.

ETS scoring specifications for the CSA were completed, approved, and checked well in advance of the receipt of student response data. These 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 computing summary data.

### Quality Control of Psychometric Processes

#### Scoring Verification

ETS developed two independent and parallel scoring structures to produce students’ scores: the Enterprise Score Key Management (eSKM) scoring system, which collected, scored, and delivered individual students’ scores to the ETS reporting system; and then the ETS PAR team computed individual student scores based on the same scoring specifications as described in subsection [*9.5.2 Development of Scoring Specifications*](#_Development_of_Scoring). The scores from the two sources were then compared for internal quality control. Any differences in the scores were discussed and resolved. All scores complied with the ETS scoring specifications and passed the parallel scoring process. This ensured the quality and accuracy of scoring and supported the transfer of scores into the Test Operations Management System (TOMS), the database of the student records scoring.

#### Psychometric Analyses

The psychometric procedures for the CSA were developed, reviewed, and approved prior to the receipt of student response data. The ETS psychometric team also developed specifications for each of the psychometric analyses performed. These specifications contain detailed descriptions of the analysis steps such as sample inclusion, analyses methods, and special handling of the data.

All psychometric analyses conducted at ETS underwent comprehensive quality checks by a team of psychometricians and data analysts. Detailed checklists and psychometric specifications were developed by members of the team for each of the statistical procedures performed on CSA results data, including item analyses, differential item functioning analyses, item response theory (IRT) calibration, equating, and scaling.

Detailed checklists were developed by members of the team for each of the statistical procedures. Classical item analyses were performed to evaluate the performance of the operational items. Classical item statistics included item difficulty and correlations between item scores and total scores. Items that were flagged for questionable statistical attributes were sent to ETS 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 fits the data. ETS then presented and reviewed the calibration results with the CDE for approval.

Once raw-to-scale-score conversion tables for each form were generated, psychometricians carried out quality control checks on each scoring table to verify

* all possible raw scores for each form were included in the tables;
* the lowest obtainable scale score and the highest obtainable scale score matched the specifications for each grade level, respectively; and
* the threshold score for the reporting range was correctly identified.

After all quality control steps were completed and any differences were resolved, one final inspection of scoring tables was conducted prior to uploading the tables to eSKM for score reporting.

### Quality Control of Reporting

To ensure the quality of CSA 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 Student Score Reports (SSRs) through ETS’ patented QC Interrogator software, which compares elements of the SSR to acceptable values to identify errors and is used in conjunction with human review to detect errors on every score report batch as part of quality control procedures
3. Proofreading of the quality control and production reports by the CDE and ETS prior to making reports available to the LEA for download in TOMS and 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 accurate codes and names. The CDE provided a revised LEA Master File to ETS throughout the year as updates became available.

After the reports were validated in accordance with the CDE’s requirements, a set of reports representing all possible grade levels, content areas, and reporting outcomes was provided to the CDE and ETS for review and approval. Electronic reports were sent on the actual report template to the CDE. The CDE and ETS reviewed and approved the reports after a thorough examination.

Upon the CDE’s approval of the reports generated for the quality control LEAs, ETS proceeded with the first batch of report production. The reviewed set of reports incorporated CDE-selected LEAs and provided the final check prior to generating all reports and making them available electronically for download in TOMS and for student information systems through an application programming interface.

#### Exclusion of Student Scores from Summary Reports

ETS provided the CDE with reporting specifications that documented when to exclude student scores from summary reports. These specifications included the logic for handling submitted assessments that, for example, indicated the student tested but responded to no items, was absent, was not tested because of parent/guardian request, or did not complete the assessment because of illness. The methods for handling other anomalies were also covered in the specifications. These anomalies are described in more detail in [*7.3.2 Special Cases*](#_Special_Cases_1).

### Quality Control of End-to-End Testing

ETS conducted end-to-end testing prior to the start of the test administration. The purpose of this testing is to verify that all systems, processes, and resources were ready for the operational administration. Once released from processing, the test results were sent through the system for scoring and reporting. SSRs were created, along with data files for subject-matter experts in the teams to review and verify.

#### Computer-based Assessments

ETS employed a number of strategies to verify ongoing systems performance, including monitoring of system availability and system usage. Time was allotted for UAT to confirm that the systems met requirements and to make identified corrections before final deployment. To accomplish system acceptance and sign-off, ETS deployed systems to a staging area, which mirrors the final production environment, for operational testing and UAT. Final approval by the CDE triggered final deployment of the system.

To begin the quality control process for end-to-end testing of the administration, the ETS program and resolutions teams prepared by entering responses in computer-based assessments for all grade spans and domains. These responses were entered for fictitious students in selected schools and across several LEAs. Each student’s test was completed with responses that were all correct, all incorrect, and combinations of correct and incorrect. These response combinations were the expected results across score ranges. The responses were sent for processing, including for system quality control of computer-based assessments.

Once released from processing, the test results were sent through the system for scoring and reporting. SSRs were created, along with data files for subject-matter experts in the teams to review and verify. Individual SSRs were generated on the basis of the fictitious students when 100 percent quality control was demonstrated by ETS’ Resolution staff.

### Reference

Educational Testing Service. (2014). *ETS Standards for Quality and Fairness*. Princeton, NJ: Educational Testing Service.

## Continuous and Systematic Improvement

The fourth operational administration of the California Spanish Assessment (CSA) occurred in 2021–22. Since the first operational administration in 2018–19, continuous efforts have been made to improve the assessments. This chapter summarizes accomplishments and ongoing improvements for the CSA as well as strategies to implement possible future improvements.

### 2021–22 Feedback for Continuous Improvement Survey

The California Assessment of Student Performance and Progress (CAASPP) program annually solicits feedback from educators through the Feedback for Continuous Improvement Survey. Local educational agency (LEA) and test site staff, as well as test administrators and test examiners, were invited to participate in the 2021–22 Feedback for Continuous Improvement Survey. Its goal was to highlight successes and identify areas for improvement. A total of 4,834 survey respondents participated in this survey for the 2021–‍22 administration, compared to 1,615 respondents for the previous year. The California Department of Education (CDE) and ETS use key recommendations from educators to implement positive changes in the following administration year.

Educators provided valuable feedback for potential improvements to the future administration of the CAASPP and the English Language Proficiency Assessments for California (ELPAC) by reporting some lessons they learned in 2021–‍22. Based on those lessons and suggestions for improvement, the *CAASPP and ELPAC Feedback for Continuous Improvement Survey and Focus Groups Report* (CDE, 2022) presents recommendations for the CDE, with the goal of enhancing the administrative support provided to LEAs and schools for future CAASPP and ELPAC test administrations. Refer also to subsection [*5.3.4 Feedback for Continuous Improvement Survey*](#_Feedback_for_Continuous_1) for assessment-specific results.

#### Outreach

In response to the LEA feedback, ETS is implementing the following improvements for the 2022–23 operational administration:

* Moving Data Entry Interface instructions to test administration manuals instead of having the content in a separate manual
* Updating regularly both the online Get Answers lookup tool as well as Anita, the chatbot, with new and frequently asked questions
* Providing Coffee Sessions on the second Tuesday of every month to give LEAs an opportunity to ask questions and receive timely updates
* Creating new demonstration videos and updating existing videos, where possible, with suggestions provided by LEAs
* Providing training on the use of accessibility resources in the classroom and how to make these resources available to students during test administration
* Providing hands-on training, in conjunction with self-paced video modules, to assist educators with how to identify, match, and use accessibility resources to meet students’ needs
* Improving the usability of the ETS-developed program websites by simplifying the display of information and offering materials more readily, when possible

#### Communications

During the 2021–22 test administration year, the CDE and ETS worked to streamline communications and provide LEAs with timely and relevant information throughout the year. One example of these efforts is the monthly ELPAC email communication, which is sent to LEA coordinators. After the results from the 2020–21 survey determined the need for a monthly CAASPP communication, that need was addressed. Monthly CAASPP emails were sent to the field during the 2021–22 test administration year, and survey respondents were asked about the helpfulness of this communication. These improvements will be continued during the 2022–23 administration.

### Accessibility Resources

The following changes will be implemented during the 2022–23 CSA administration:

* The color contrast designated support will include the following additional options:
* Yellow font on a black background
* Red font on a white background
* White font on a red background

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.

### Reference

California Department of Education. (2022). *2021–22 CAASPP and ELPAC feedback for continuous improvement survey and focus groups report*. [Unpublished manuscript]. Sacramento, CA: California Department of Education.

1. Data for 2021–22 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 versions of the Accessibility Matrix and the *Usability, Accessibility, and Accommodations Guidelines* that were available during the 2021–22 CAASPP administration. [↑](#footnote-ref-4)
4. For students in grades six through eight and high school [↑](#footnote-ref-5)
5. The language standards, which focus on vocabulary, can be seen as an integral support of each of the four skills. [↑](#footnote-ref-6)